# Collinear Unpolarised Fragmentation Functions:

## a general overview and an (incomplete) collection of recent and preliminary results


Valerio Bertone

NIKHEF and VU Amsterdam

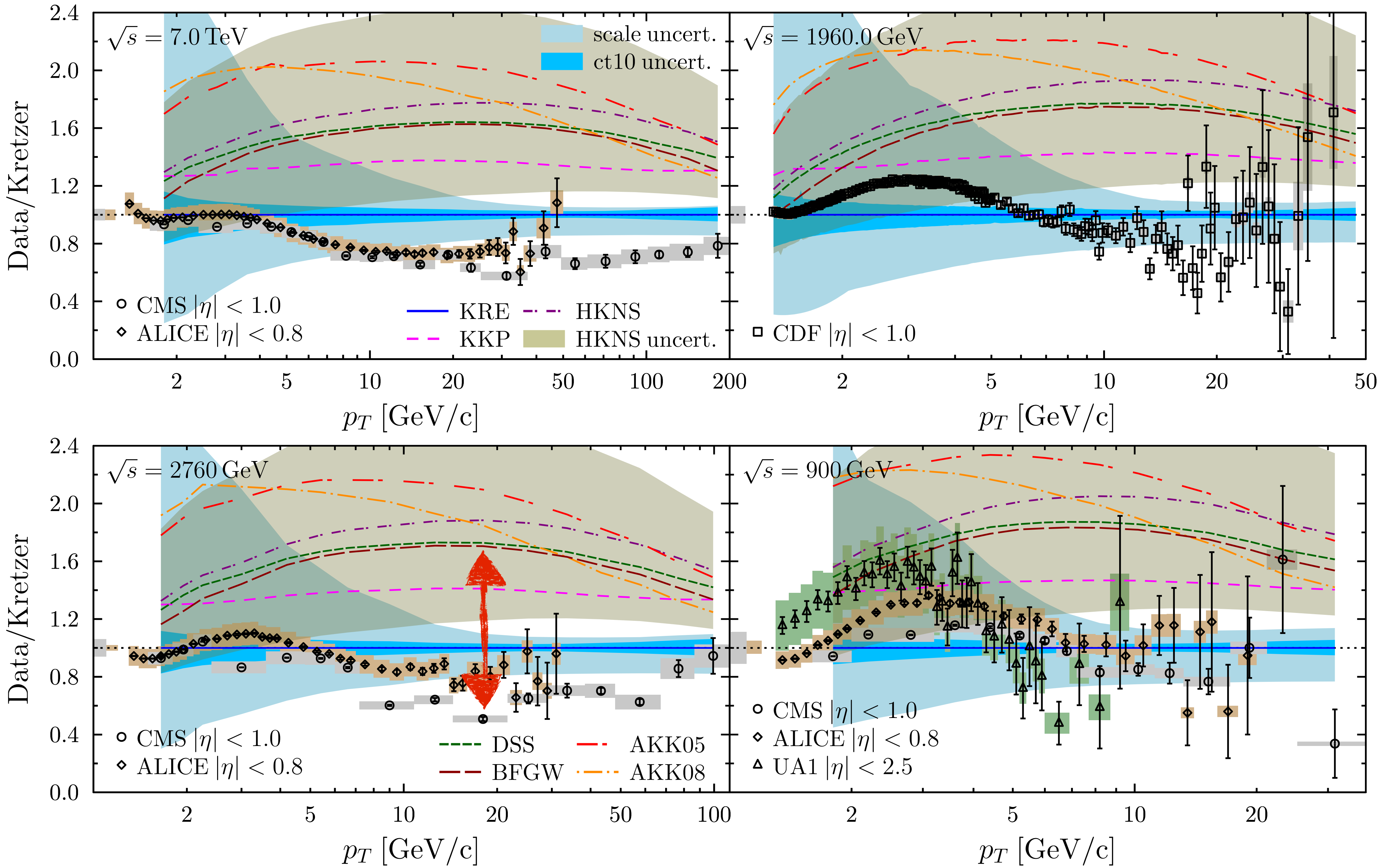

Workshop on Fragmentation Functions

February 20, 2018, Stresa (Italy)

# Interesting facts

*Why bother about fragmentation functions*

Comparison data/theory for inclusive **charged-hadron $p_T$ spectra**:




d'Enterria *et al.* [ArXiv:1311.1415]

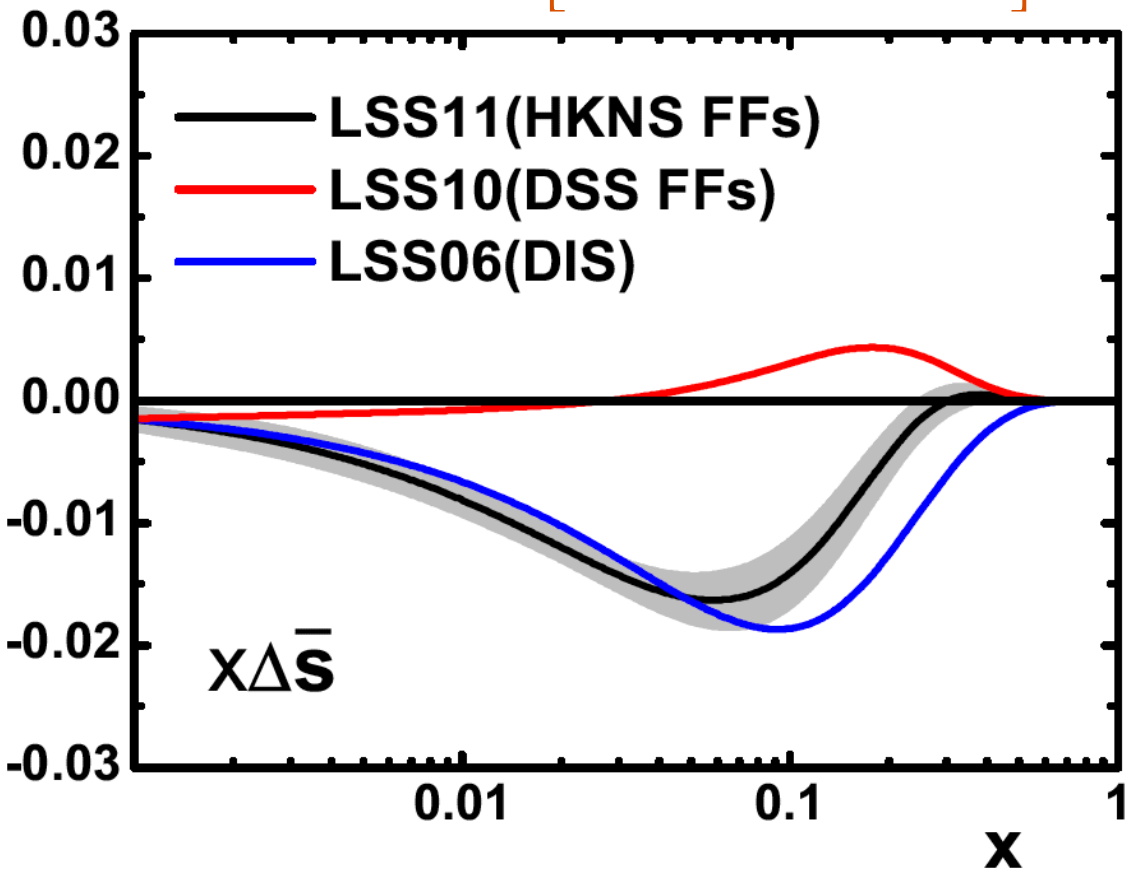
Large energy data tend to be **overshot** by predictions obtained with most of the current FF sets ⇒ too hard gluon FF at large $z$?

# Interesting facts

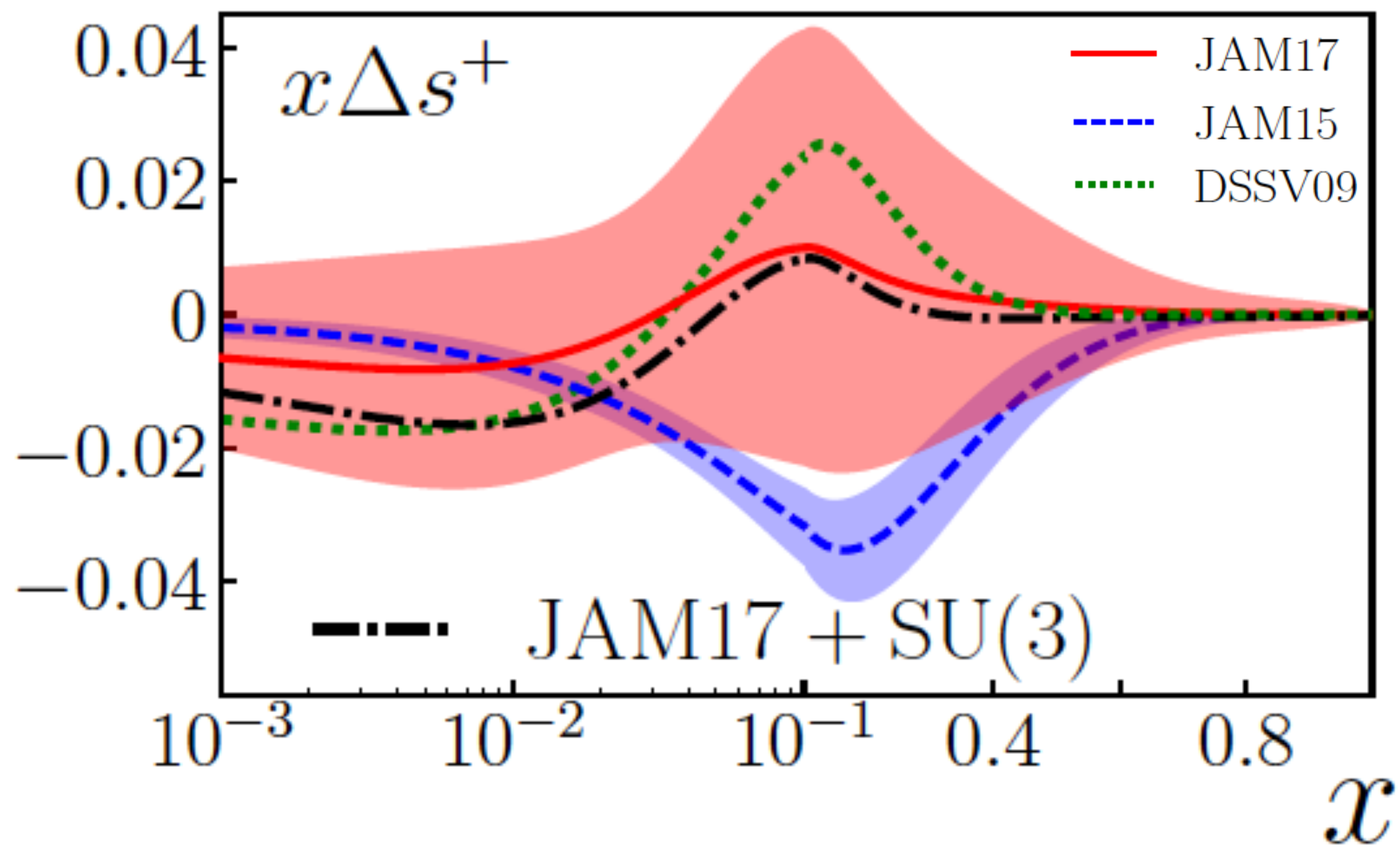
## *Why bother about fragmentation functions*

Extraction of the **longitudinally polarised parton distribution functions**:

Laeder *et al.* [ArXiv:1103.5979]




J. J. Ethier *et al.* [ArXiv:1705.05889]



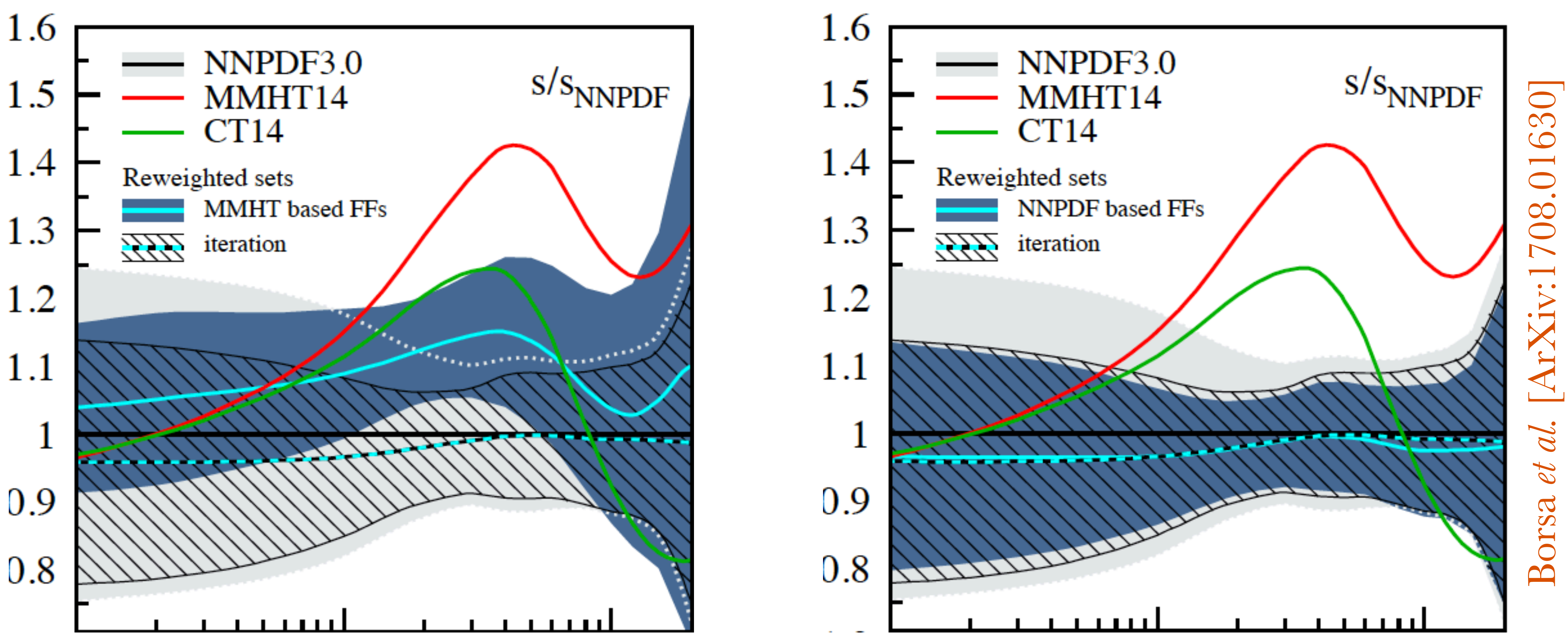

- In the presence of semi-inclusive DIS (SIDIS) data the **strange quark distribution** is very sensitive to the choice of the FF set used in the analysis.
- Even fitting PDFs and FFs simultaneously does no lead to a definitive answer.

# Interesting facts

*Why bother about fragmentation functions*

- **SIDIS** multiplicities depend on **PDFs** and thus the precise data from HERMES and COMPASS should in principle help constrain PDFs (noticeably the **strange** PDFs if a kaon is produced in the final state).

- On the other hand, SIDIS multiplicities also depend on **FFs** whose determination in turn often depends on PDFs.




- A **combined** extraction of PDFs and FFs is the only way to overcome this limitation.

# Everything starts from…

The **collinear factorisation theorem** (assumed to work):

$$d\sigma_{\rm had} = \boxed{W_{\{i\}}} \otimes \boxed{\mathcal{L}_{\{i\}}} d\Phi$$

**Hard cross sections**:

- process dependent,
- high-energy dominated,
- computable in perturbation theory.

**PDFs and/or FFs**:

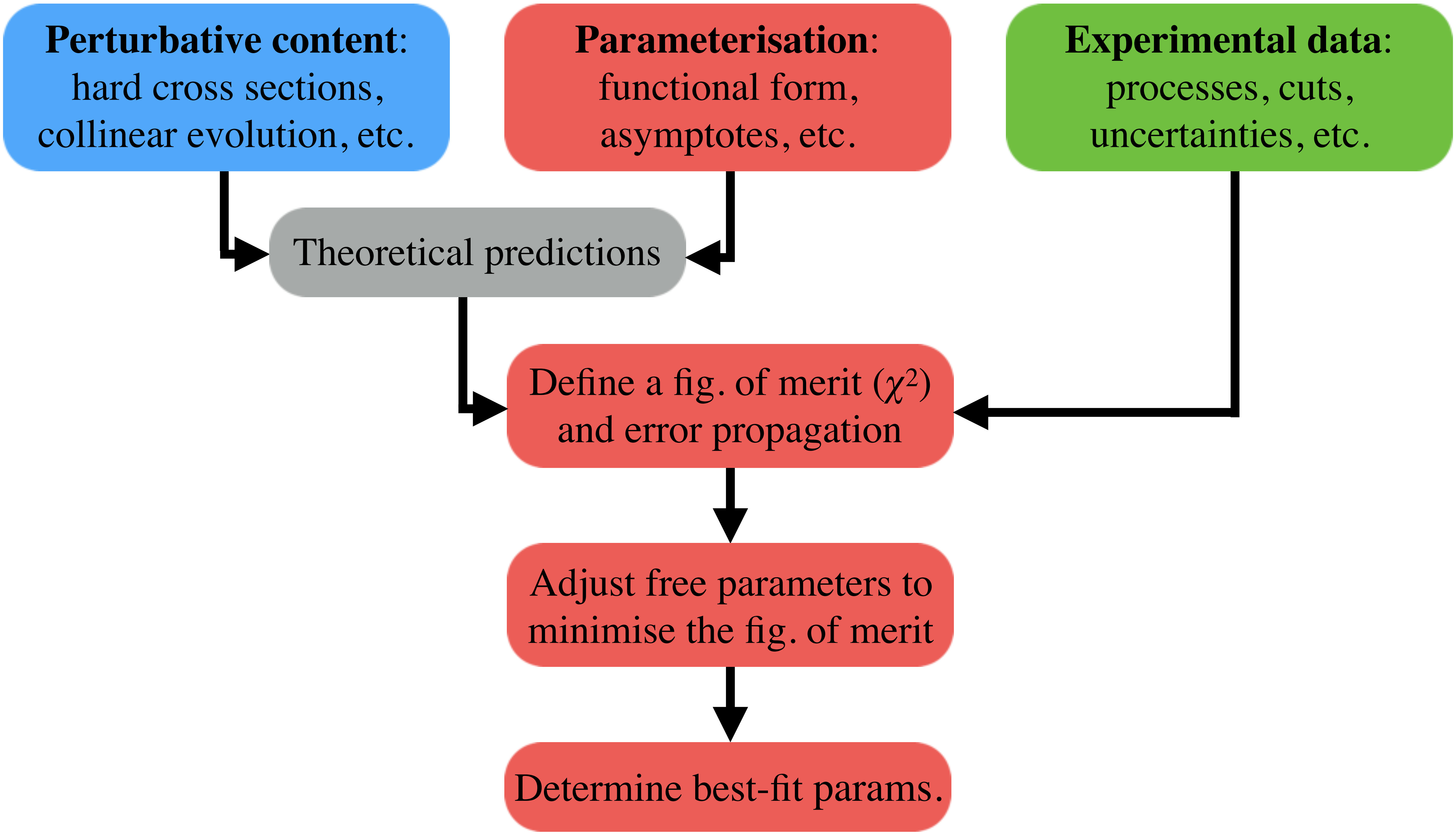
- universal (given the hadronic species),
- low-energy dominated,
- perturbation theory inapplicable.

How do we determine PDFs and FFs?

$\Downarrow$

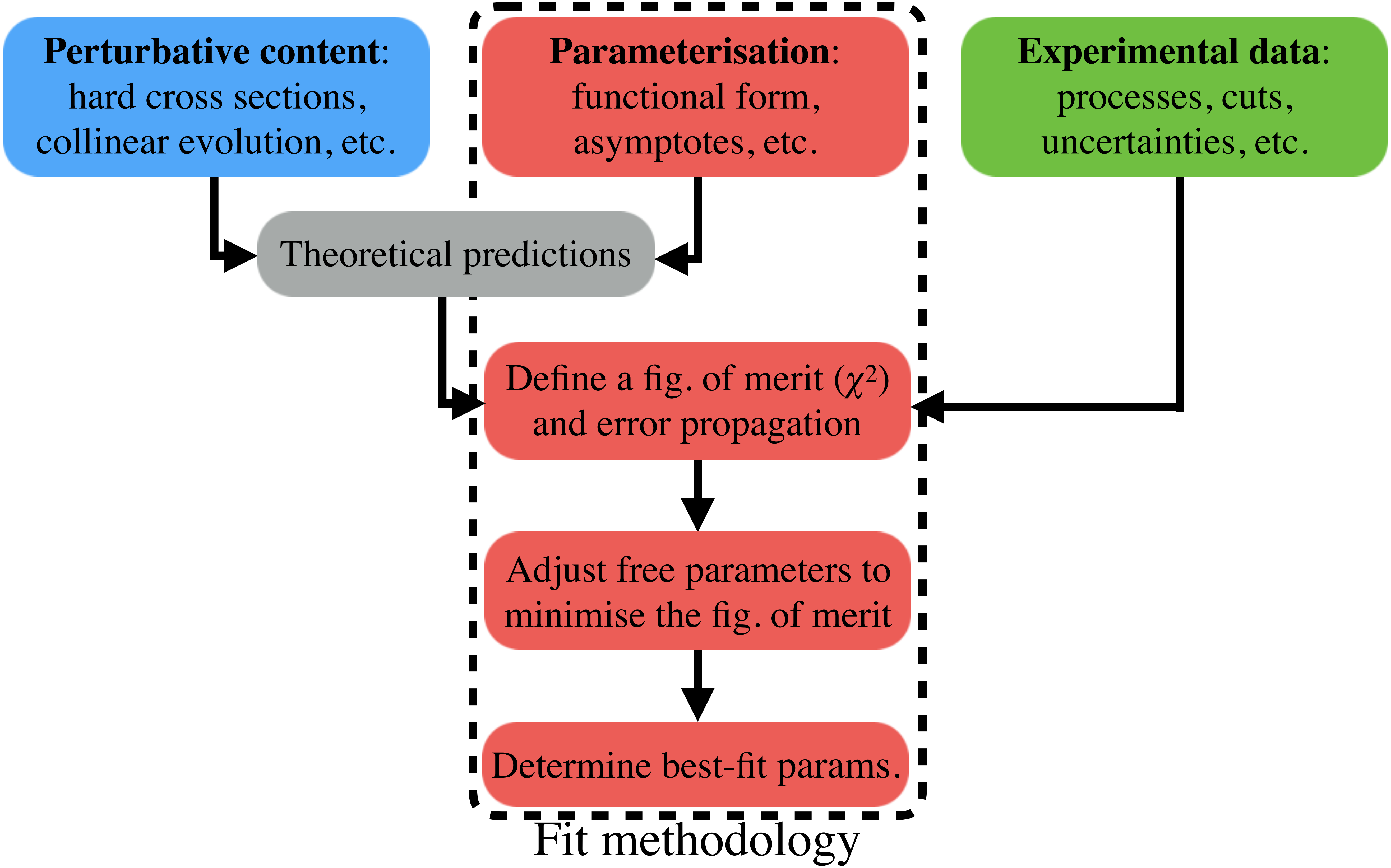
Currently, the most accurate and reliable way is through **fits to data**.

# The general strategy




Each box requires a choice. **Different choices** lead to **different determinations**.

# The general strategy



# Fit methodologies

## *Parameterisation: the "standard" approach*

- Distributions are parametrised by means of the function form:

$$f_i(x) = A_i\, x^{\alpha_i} (1-x)^{\beta_i} P_i(x)$$

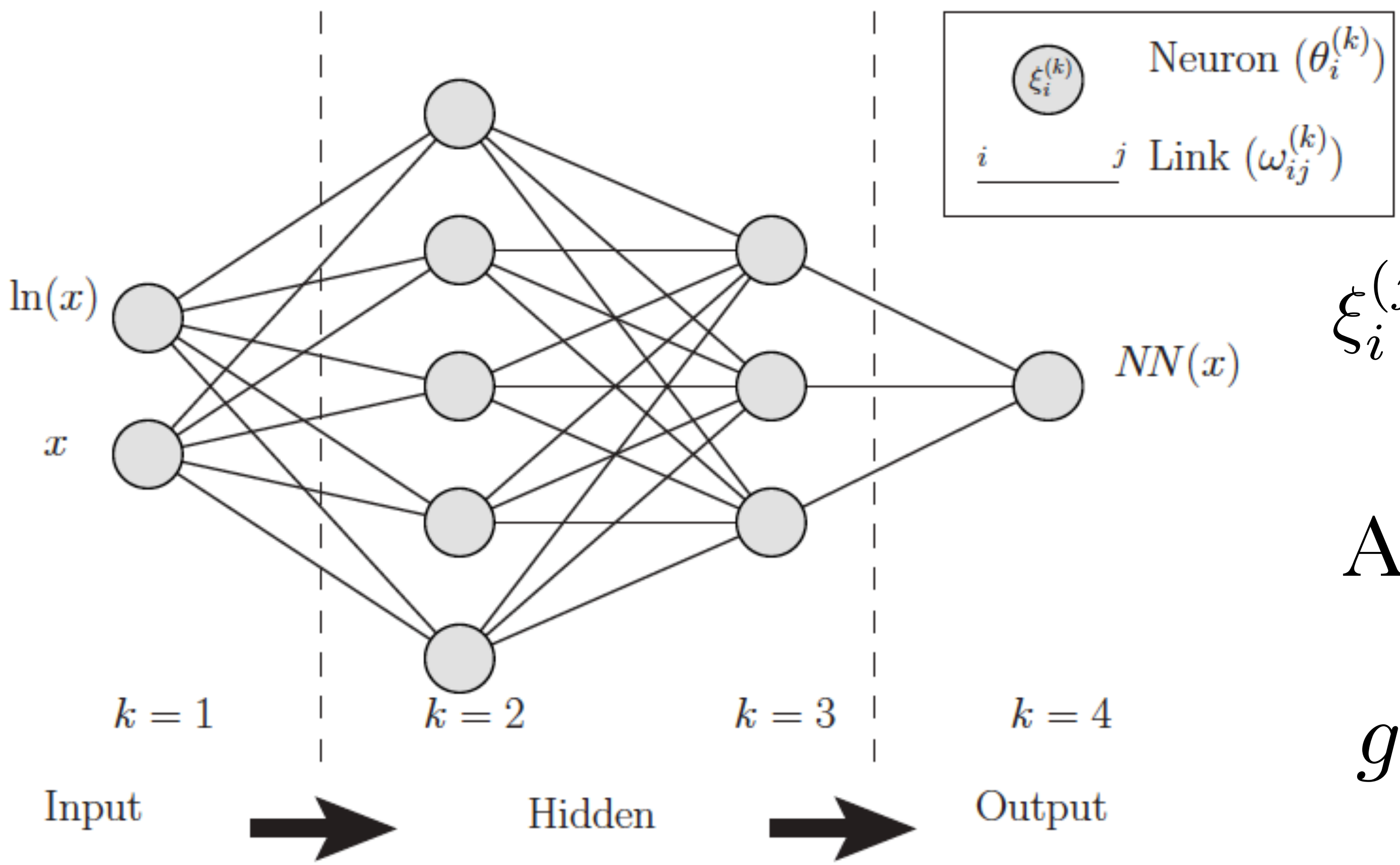
with:

$$P_i(x) = \begin{cases} 1 \\ 1 + \gamma_i x \\ 1 + \gamma_i x + \delta_i \sqrt{x} \\ \dots \end{cases}$$

- **O(3-5) free parameters** for each distribution.
- **Asymptotic behaviour** defined by the exponents $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_i}$.
- Easy to transform analytically in **Mellin space**.
- **Easy to handle** in a fit thanks to its simplicity.
- Potential **source of bias**.

# Fit methodologies
## *Parameterisation: neural networks (in NNFF1.0)*

- Distributions are parametrised in terms of NNs with arch. (2-5-3-1):




$$\xi_i^{(j)} = g\left(\sum_k^{\text{(j-1)th layer}} \xi_k^{(j-1)} \omega_{ki}^{(j)} - \theta_i^{(j)}\right)$$

Activation function:

$$g(x) = \mathrm{sign}(x)\ln(|x|+1)$$

- Each NN has **37 free** parameters each.
- Distributions are expressed as $f_i(x) = \mathrm{NN}_i(x) - \mathrm{NN}_i(1)$
  - The $\mathrm{NN}_i(1)$ term ensures that $f_i(x) \xrightarrow[x\to 1]{} 0$
- NNs are **flexible** and thus limit biases but are **harder to handle**.

# Fit methodologies

## *Figure of merit: the $\chi^2$ definition*

- A crucial aspect in the determination of PDFs/FFs is the definition of the **figure of merit** to be minimised/maximised.

- A popular choice is the $\chi^2$ but **many variants** are possible:
  - No correlation, no normalisation unc.: $$\chi^2 = \sum_{i=1}^{N_{\rm dat}} \frac{(T_i - D_i)^2}{\sigma_i^2}$$
  - No correlation, with normalisation unc.: $$\chi^2 = \sum_{j=1}^{N_{\rm exp}} \left[ \left( \frac{1-\mathcal{N}_j}{\delta\mathcal{N}_j} \right) + \sum_{i=1}^{N_{\rm dat}^j} \frac{(\mathcal{N}_j T_i - D_i)^2}{\sigma_i^2} \right]$$
  - Nuisance parameters: $$\chi^2 = \sum_i \frac{\left[ T_i \left( 1 - \sum_j \gamma_j^i b_j \right) - D_i \right]^2}{\delta_{i,\rm unc}^2 T_i^2 + \delta_{i,\rm stat}^2 D_i T_i} + \sum_j b_j^2$$
  - Covariance matrix: $$\chi^2 = \sum_{ij} (T_i - D_i) \sigma_{ij}^{-2} (T_j - D_j)$$

- Due to the **D'Agostini bias**, a sound treatment of normalisation uncertainties requires particular care (*e.g.* the $t_0$ prescription).

# Fit methodologies
## *Error propagation*

- A faithful determination implies an estimate of the **uncertainty** on FFs/PDFs propagating from the **experimental** dataset.

1. **Hessian** method: the $\chi^2$ is **expanded** around its minimum $\mathbf{a}_0$:

$$\chi^2(\{\mathbf{a}\}) \simeq \chi^2(\{\mathbf{a}_0\}) + \underbrace{\frac{1}{2} \left. \frac{\partial \chi^2}{\partial a_i \partial a_j} \right|_{\mathbf{a}_0}}_{H_{ij}} (a_i - a_{0i})(a_j - a_{0j})$$

The Hessian matrix $H_{ij}$ is **diagonalised** and an uncertainty along each eigenvector is defined as $\Delta\chi^2 = 1$ (sometimes a **tolerance** is introduced).

2. **Monte Carlo** sampling: artificial **replicas** of the dataset generated as:

$$D_i^{(k)} = D_i + r_i^{(k)} \sigma_i \,, \qquad \begin{array}{l} k = 1, \ldots, N_{\rm rep} \\ i = 1, \ldots, N_{\rm dat} \end{array}$$

$r_i^{(k)}$ is a *normally distributed* and *univariate* random number. A fit is performed to each replica to produce $N_{\rm rep}$ sets of distributions $\{f_k\}$, such that:

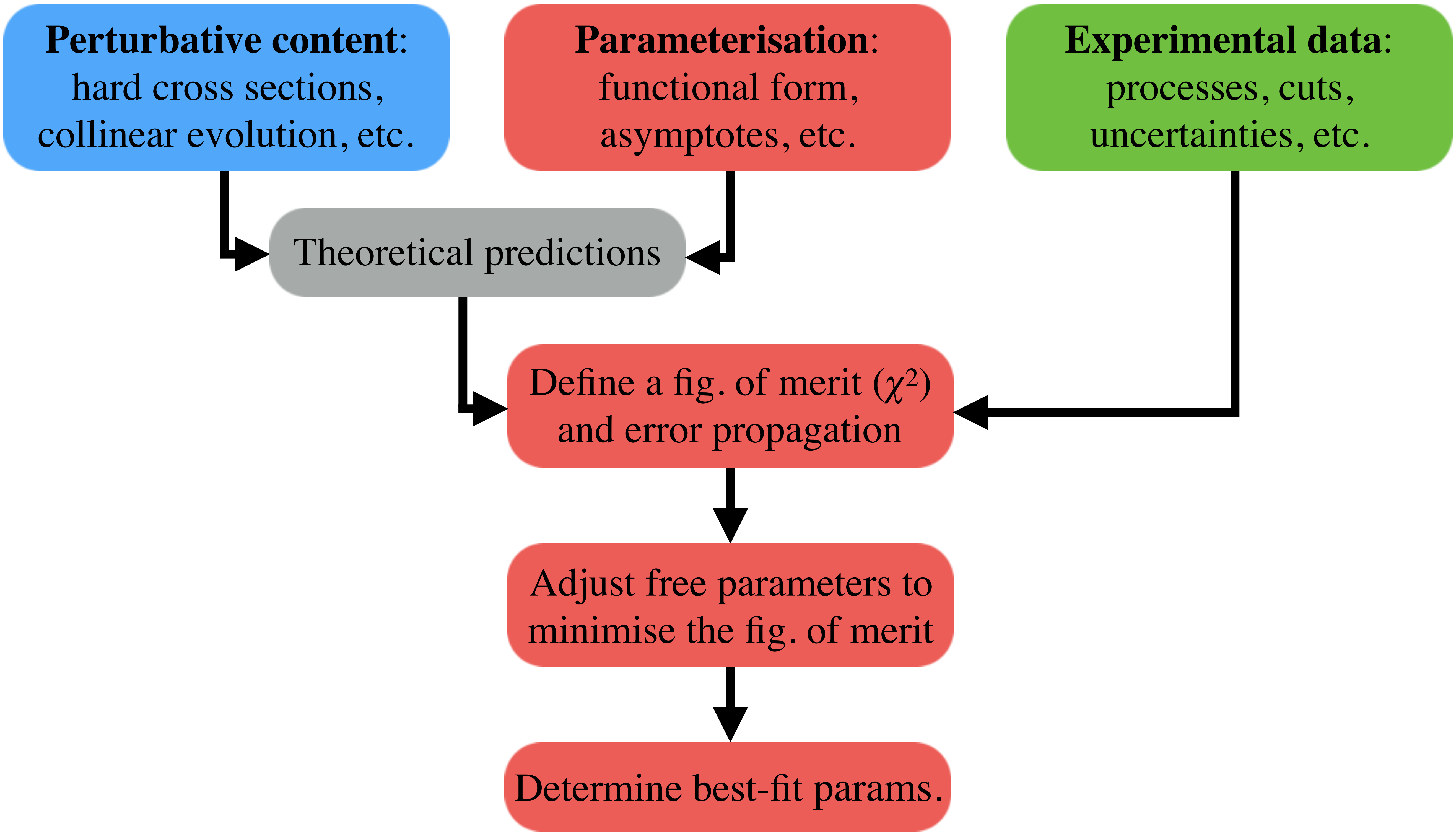
$$\langle \mathcal{O} \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{O}[f_k] \quad \text{and} \quad \sigma_{\mathcal{O}} = \sqrt{\langle \mathcal{O}^2 \rangle - \langle \mathcal{O} \rangle^2}$$

# Fit methodologies

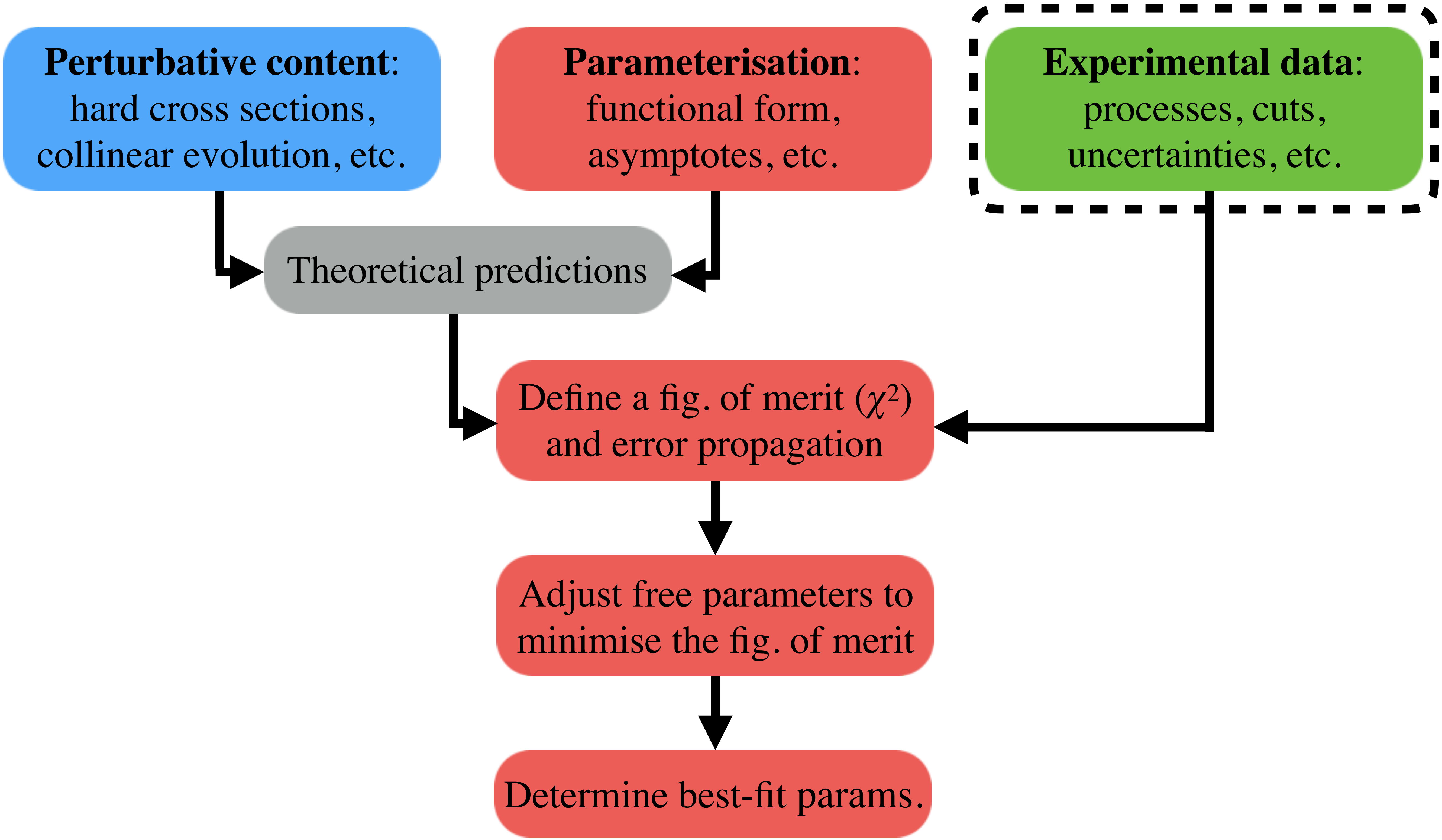
## *Minimisation and stopping*

- Simple parameterisations (**O(20) free parameters**) are usually fitted using **MINUIT** (or similar):
  - the absolute minimum of the $\chi^2$ is found *deterministically* by computing (numerically or analytically) the first derivative and moving **downhill**.
- A NN parameterisation (**O(200) free parameters**) generates a too complex parameter space to be treated deterministically:
  - a **genetic algorithm** is typically used to explore the parameter space,
  - this avoids getting trapped into **local minima** of the $\chi^2$.
- The extreme flexibility of NNs may cause **overfitting**, *i.e.* statistical fluctuations of the data sample may be unwillingly fitted:
  - the **cross-validation** method allows one to overcome this problem.
- More refined algorithms based on **machine-learning** techniques are currently being explored (*e.g.* CMA-ES).

# The general strategy

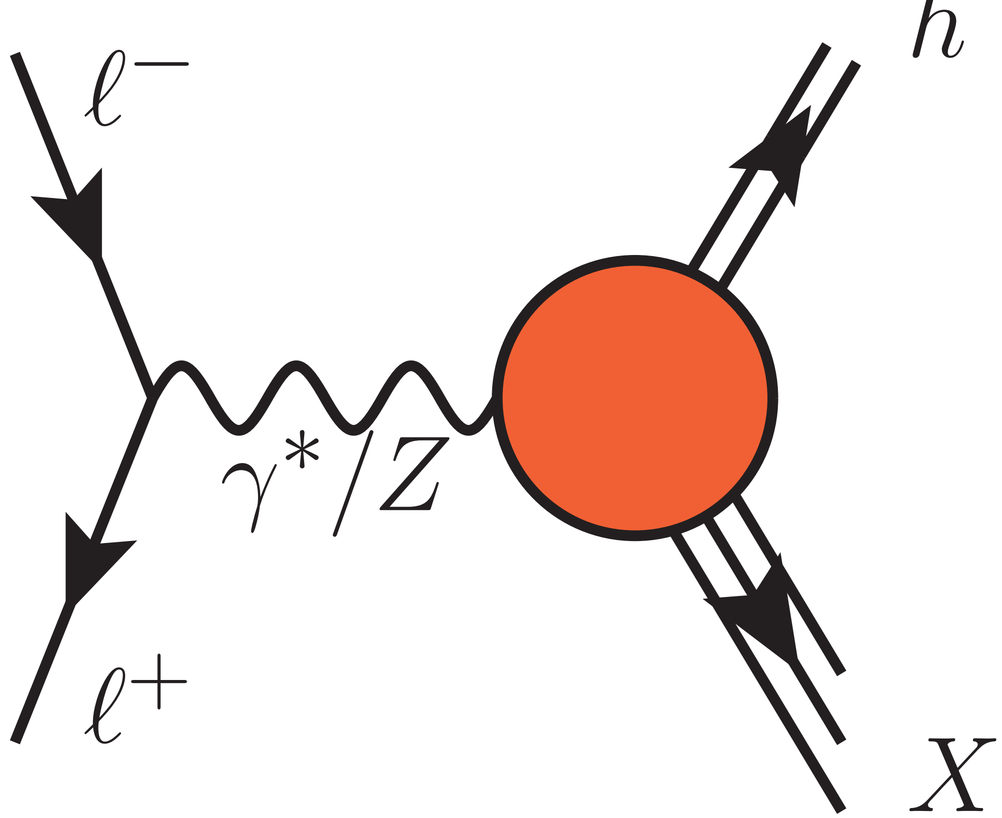


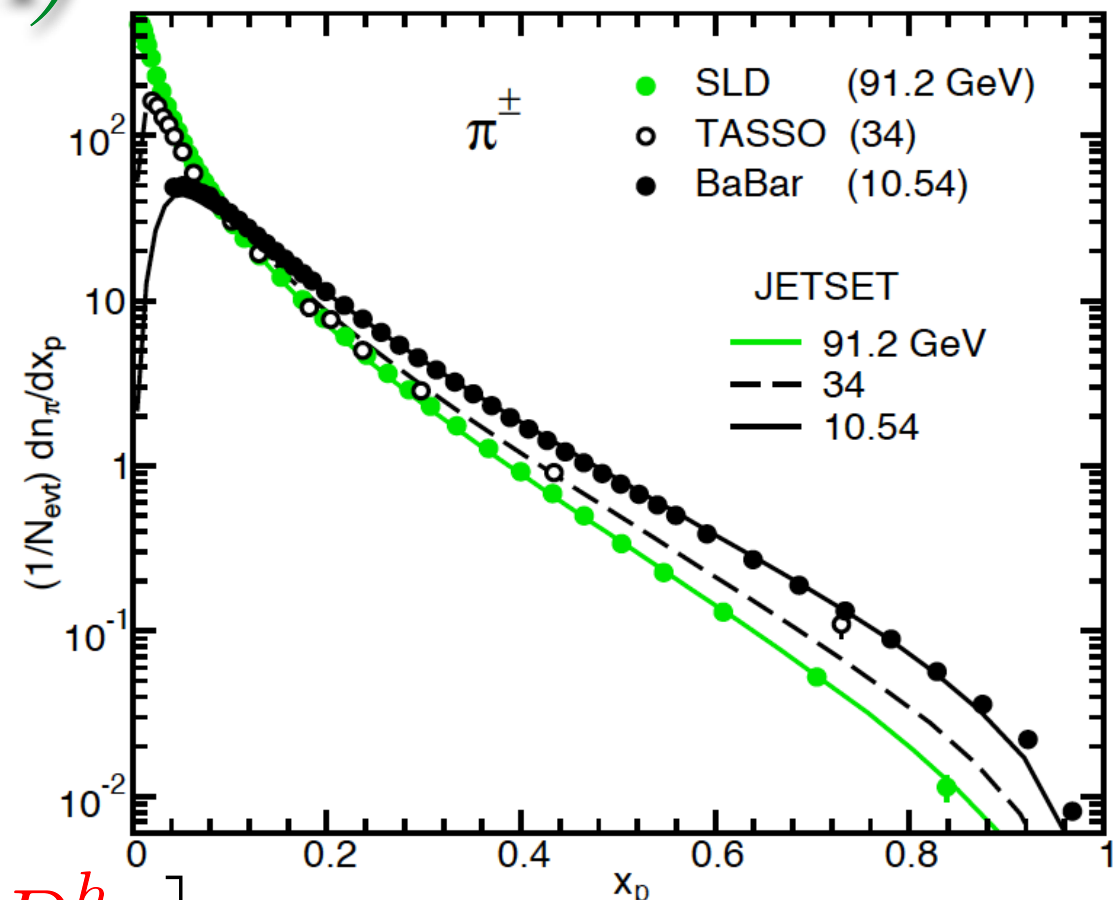
# The general strategy



# Experimental data

*Single Inclusive Annihilation (SIA)*




$$\frac{d\sigma^h}{dz} = \hat{\sigma}_0^h \left[ C_q \otimes D_\Sigma^h + C_g \otimes D_g^h + C_{\rm NS} \otimes D_{\rm NS}^h \right]$$

- ✅ **Clean** channel: only FFs involved,
- ✅ **higher-order** corrections to NNLO,
- ✅ **precise data** available (BELLE/BABAR).
- ❌ **No flavour separation**,
  - ✅ tagged data for heavy-quark FFs.
- ❌ gluon distribution **suppressed** by $\alpha_s$.

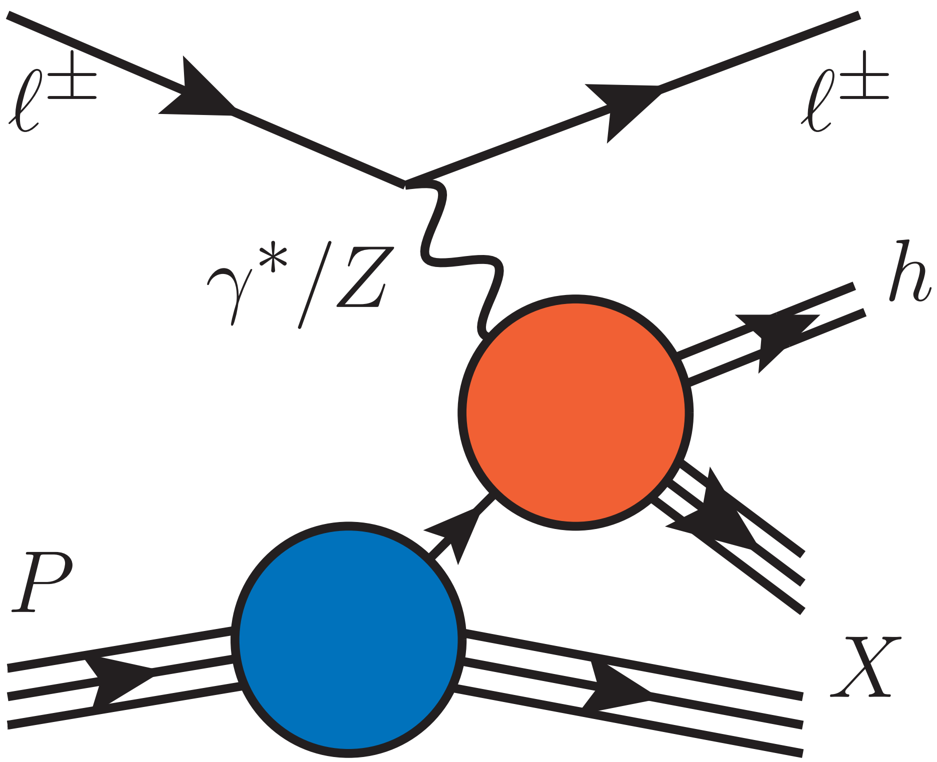
$$D_\Sigma^h = \sum_q \left( D_q^h + D_{\bar{q}}^h \right) = \sum_q D_{q^+}^h$$

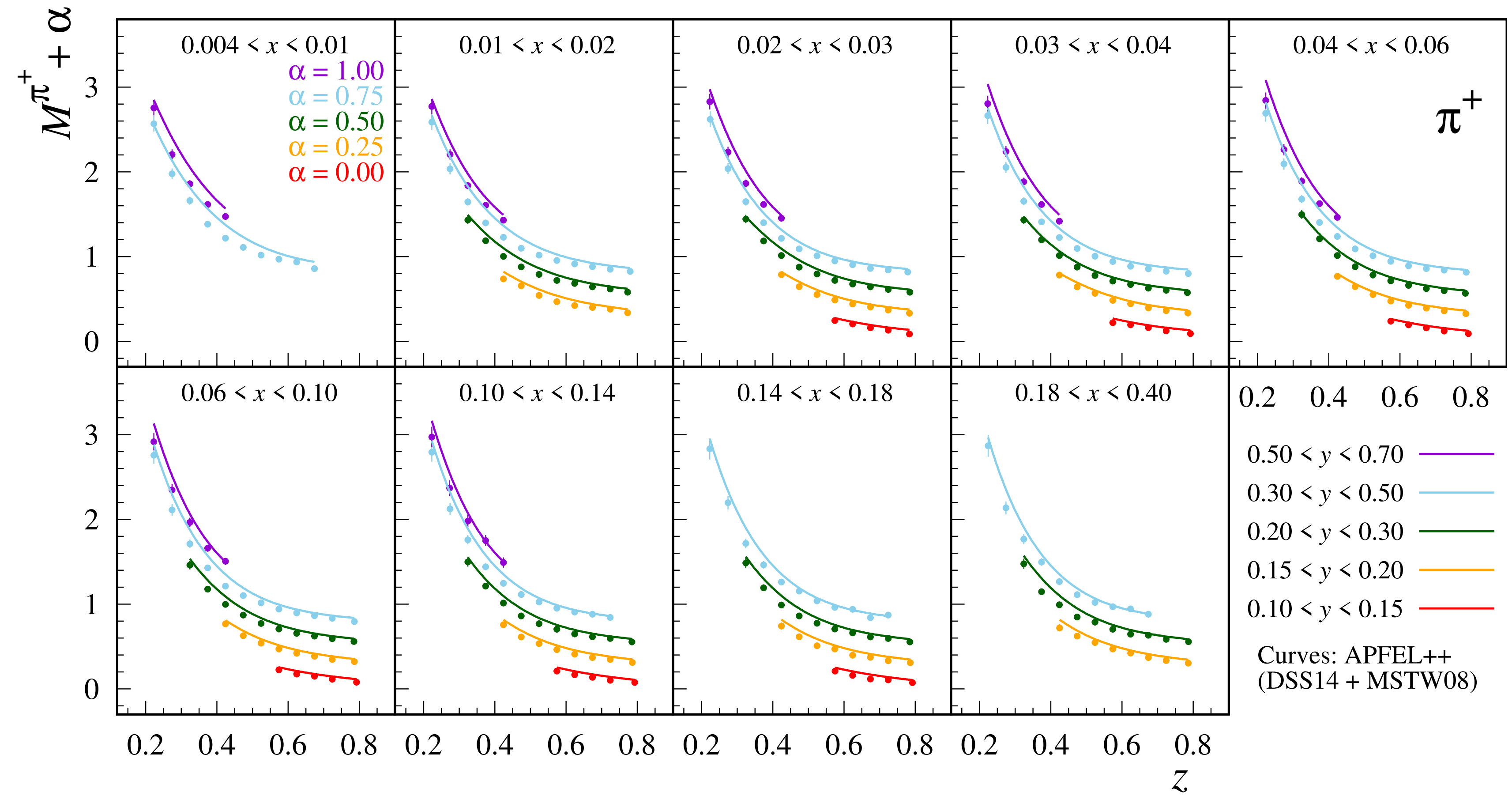
$$D_{\rm NS}^h = \sum_q \left( \frac{\hat{e}_q^2}{\langle \hat{e}_q^2 \rangle} - 1 \right) D_{q^+}^h$$

$$C_q, C_{\rm NS} \propto \mathcal{O}(1) \quad \text{while} \quad C_g \propto \mathcal{O}(\alpha_s)$$

# Experimental data

## *Semi Inclusive Deep Inelastic Scattering (SIDIS)*



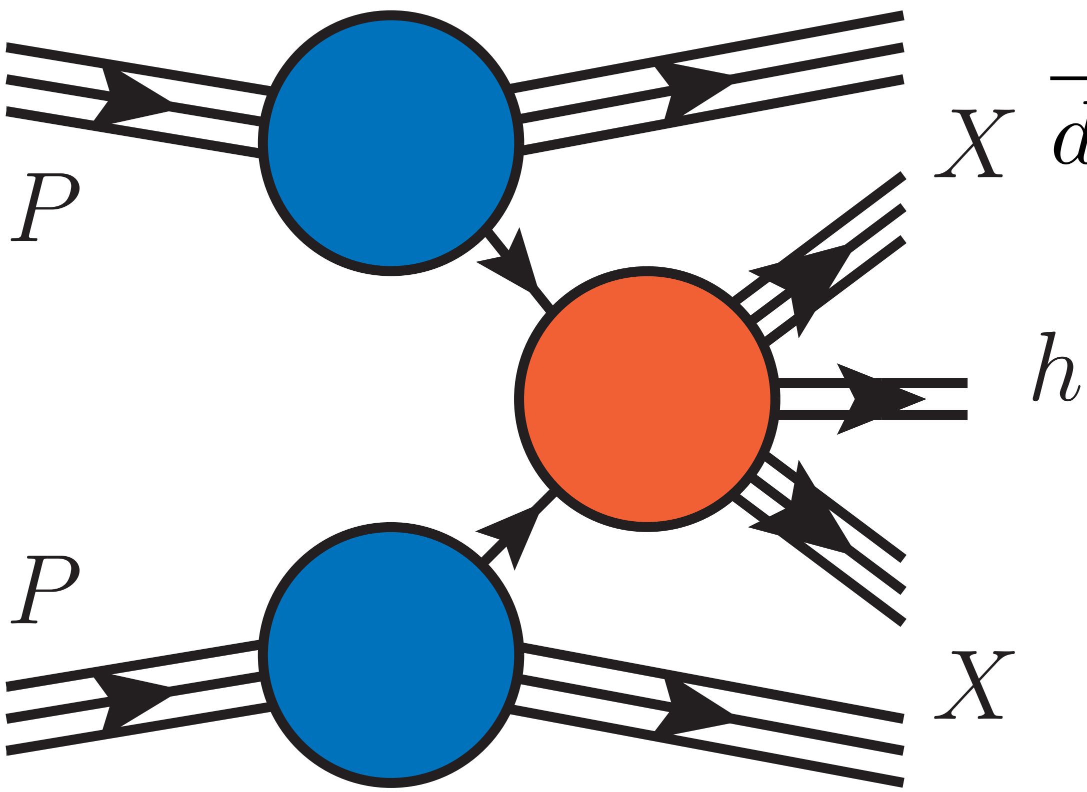




$$\frac{d\sigma^h}{dxdydz} = \hat{\sigma}_0^h \sum_{q,\bar{q}} e_q^2 \left[ f_q \otimes C_{qq} \otimes D_q^h + f_g \otimes C_{gq} \otimes D_q^h + f_q \otimes C_{qg} \otimes D_g^h \right]$$

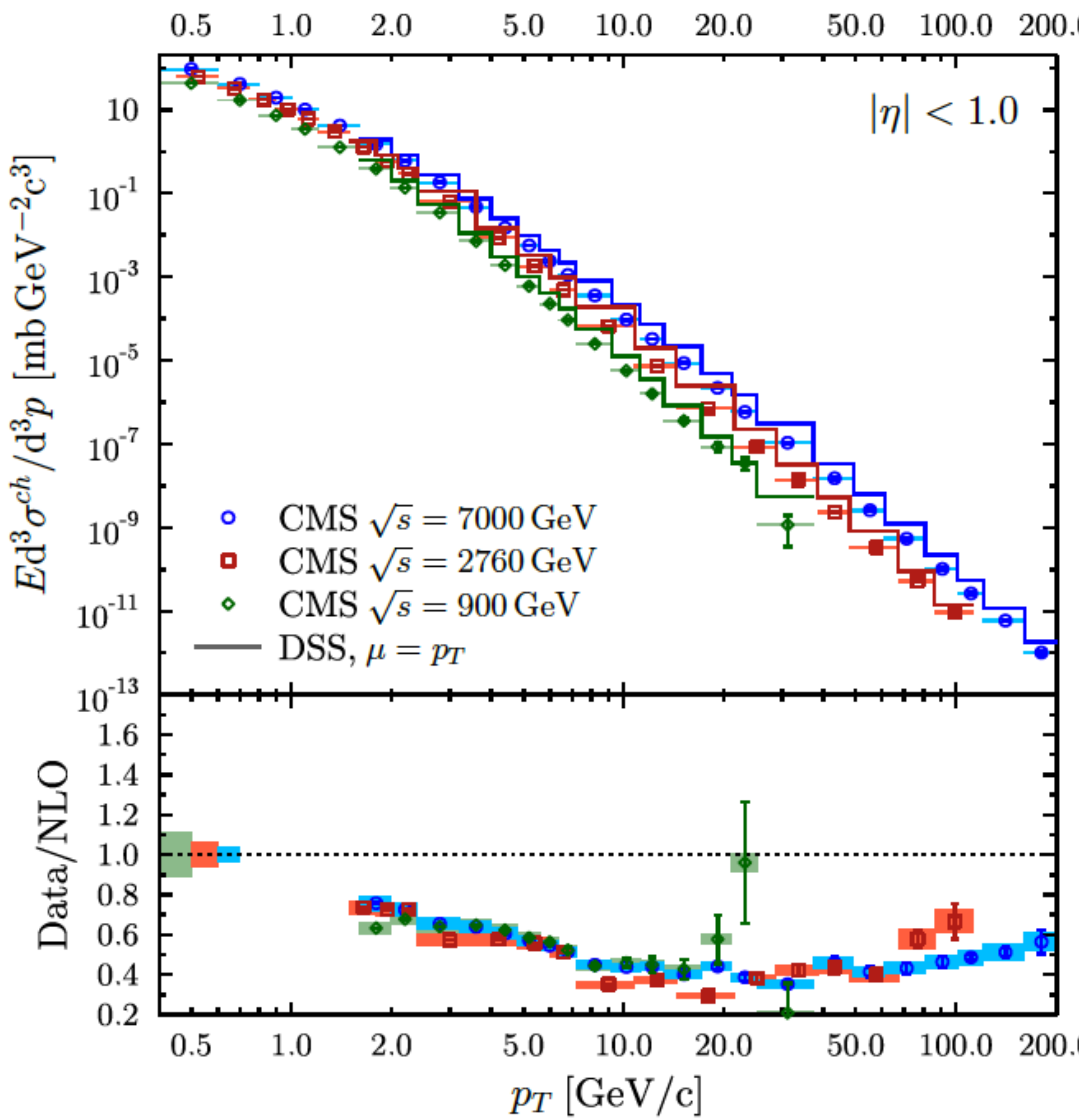
- ✅ handle on **flavour separation**,
- ✅ **precise data** available (HERMES/COMPASS).
- ❌ Involves both **FFs** and **PDFs**,
- ❌ Fully known so far up to $O(\alpha_s)$, *i.e.* NLO.

# Experimental data

*Hadroproduction in proton-proton collisions (pp)*

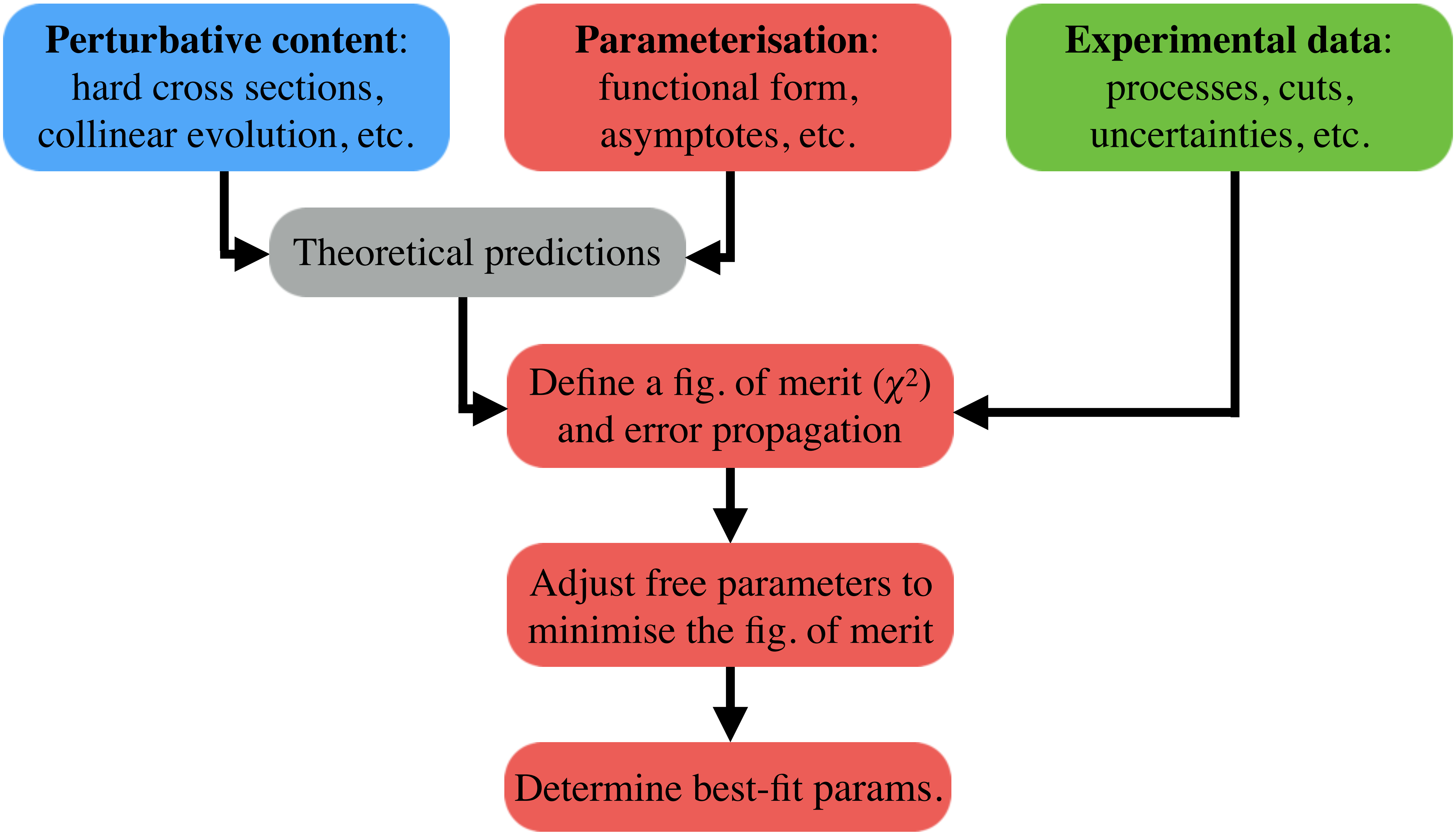



$$\frac{d\sigma^h}{dp_T d\eta} = \sum_{ijk} f_i^{(1)} \otimes f_j^{(2)} \otimes \frac{d\hat{\sigma}_{ijk}}{dp_T d\eta} \otimes D_k^h$$

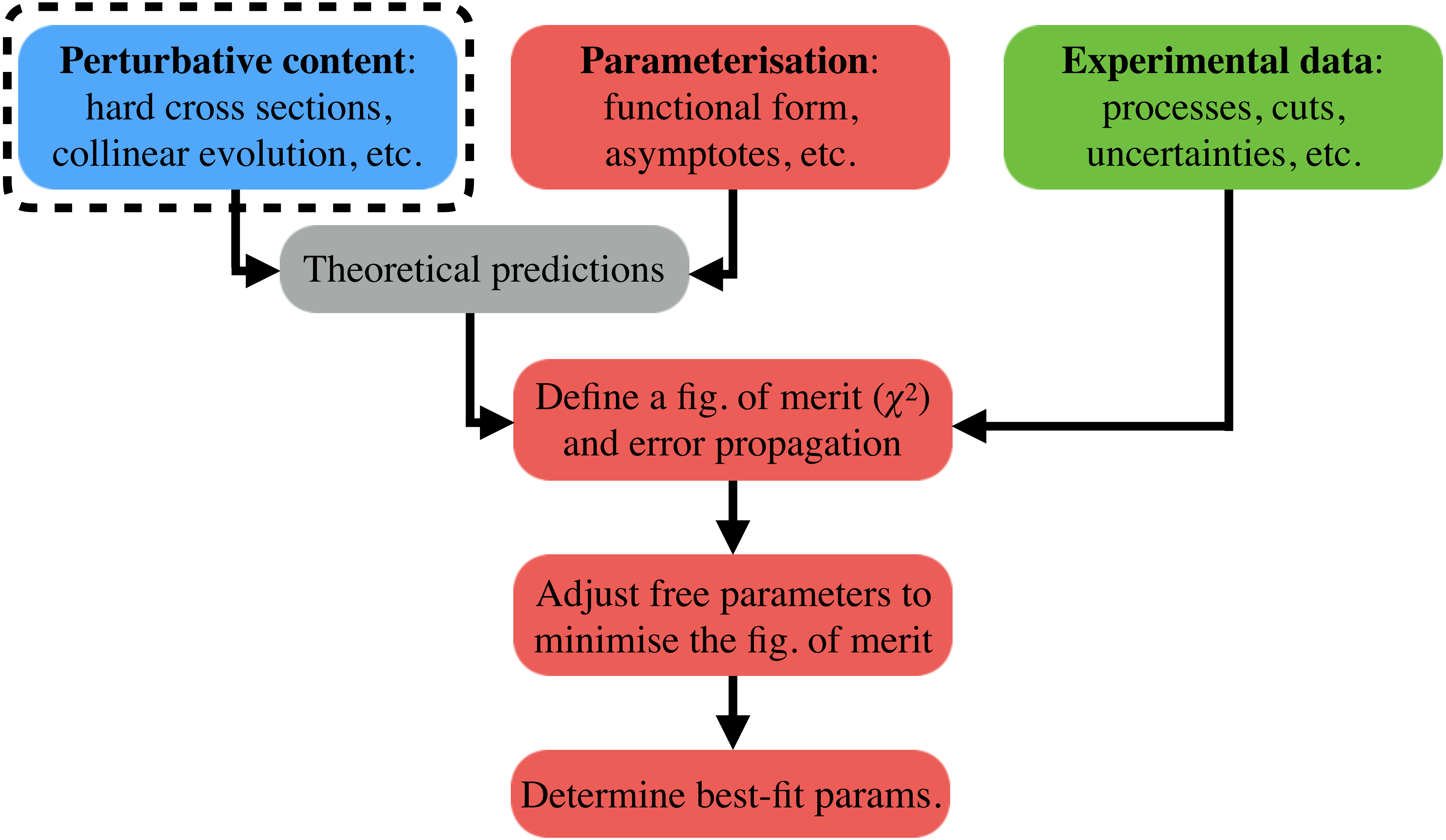



- ✓ Direct sensitivity to the **gluon FF**,
- ✓ **Scale scan** ($\mu_F \propto p_T$),
- ✓ Precise data from LHC/Tevatron
- ✗ Involves both **FFs** and **PDFs**,
- ✗ Known so far up to NLO,
  - ✗ large scale variations at low $p_T$,
- ✗ cumbersome to compute.

# The general strategy



# The general strategy



# Perturbative content
## *DGLAP evolution*

- FFs obey the standard collinear DGLAP evolution equations:

$$\mu^2 \frac{\partial}{\partial \mu^2} D^h_{\rm NS} = P_{\rm NS} \otimes D^h_{\rm NS}$$

$$\mu^2 \frac{\partial}{\partial \mu^2} \begin{pmatrix} D^h_\Sigma \\ D^h_g \end{pmatrix} = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} D^h_\Sigma \\ D^h_g \end{pmatrix}$$

- Time-like splitting functions known up to NNLO.
  A. Mitov and S. O. Moch [hep-ph/0604160], M. Gluck, E. Reya, and A.Vogt [Phys.Rev. D48 (1993)]
- Numerical implementation in the **APFEL** code:
  V. Bertone, C. Carrazza, J. Rojo [arXiv:1310.1394]
  - careful benchmark against in the the $N$-space **MELA** code,
    V. Bertone, S. Carrazza, E. R. Nocera [arXiv:1501.00494]
  - perfect agreement with **QCDNUM** (after a correction of a bug in the latter).
    M. Botje [arXiv:1602.08383]

# Perturbative content

## *Hard cross sections*

- **Single-inclusive annihilation**:
  - currently known up to O($\alpha_s^2$), *i.e.* NNLO, in the **zero-mass** scheme.
    A. Mitov and S. O. Moch [hep-ph/0604160]
  - **Mass corrections** known up to to O($\alpha_s$),
    T. Kneesch [desy-thesis-10-049]
  - **Small-*z* resummation** corrections up to NNLL,
    D. Anderle [arXiv:1611.03371]
  - **Hadron-mass** corrections (particular relevant for kaons and protons).
    e.g. A. Accardi [arXiv:1411.3649]
  - **Threshold (large-*z*) resummation** corrections up to N$^3$LL.
    S. O. Moch and A. Vogt [arXiv:0908.2746]
- **Semi-Inclusive Deep-Inelastic-Scattering**:
  - currently fully know up to O($\alpha_s$), *i.e.* NLO, in the **zero-mass** scheme.
  - partial knowledge of the O($\alpha_s^2$) corrections to $F_L$.
    D. Anderle [arXiv:1612.01293]
  - **Threshold resummation** corrections to NLL.
    e.g. D. Anderle [arXiv:1304.1373]
- **Hadroproduction in proton-proton collisions**:
  - currently fully know up to O($\alpha_s^3$), *i.e.* NLO, in the **zero-mass** scheme.
    P. Aurenche *et al.* [hep-ph/9910252]

# A selection of recent results

# Most recent determinations

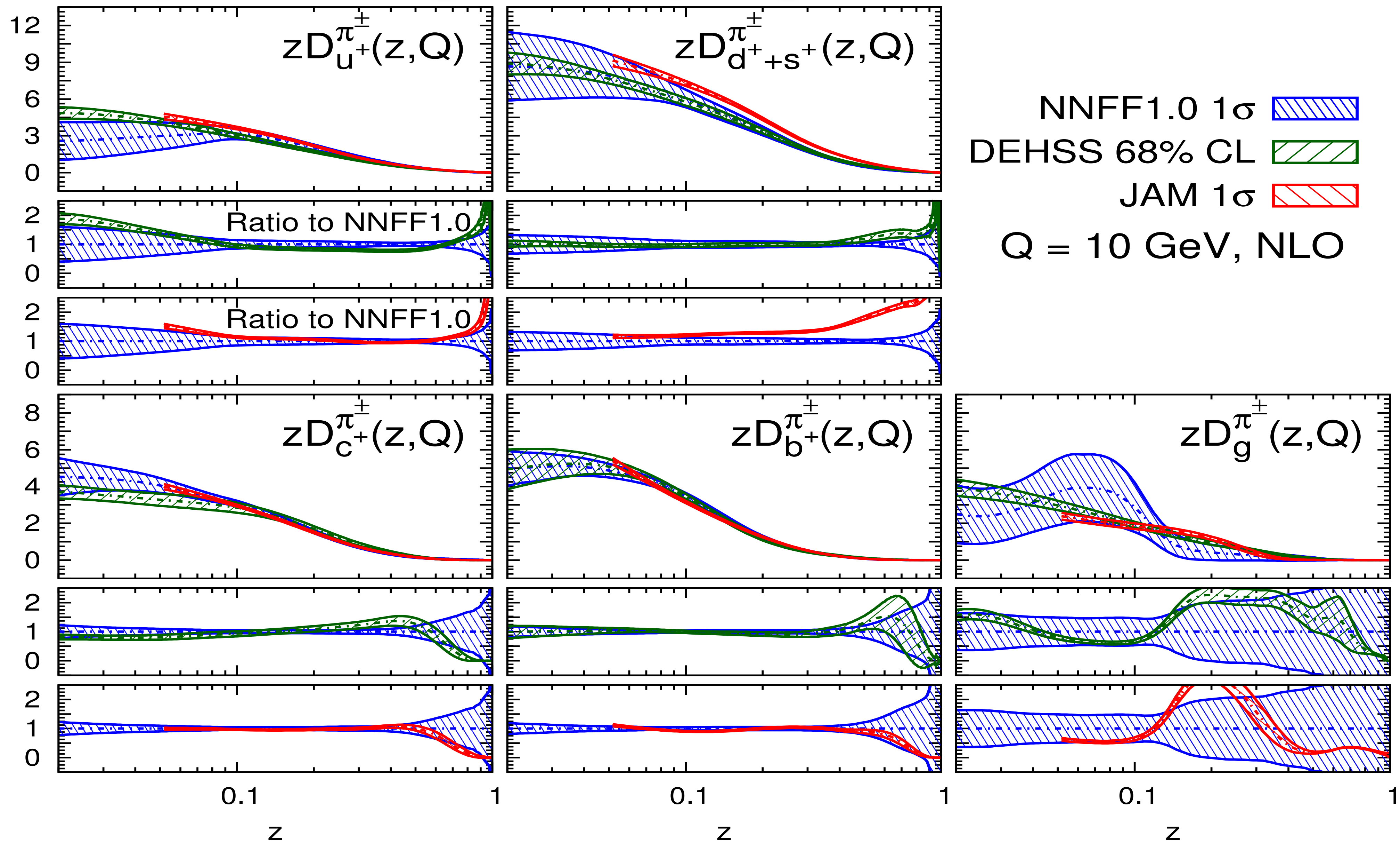
*"Global" fits: Overview*

| | DEHSS [arXiv:1410.6027]… | HKKS [arXiv:1608.04067] | JAM [arXiv:1609.00899] | NNFF1.0 [arXiv:1706.07049] |
|---|---|---|---|---|
| **Parameterisation** | standard | standard | standard | neural nets |
| **Figure of merit** | $\chi^2$, only norm. | $\chi^2$, only norm. | $\chi^2$, nuisance pars. | $\chi^2$, cov. matrix |
| **Error propagation** | Hessian | Hessian | Monte Carlo | Monte Carlo |
| **Minimisation** | MINUIT | MINUIT | MINPACK | GA/CMA-ES |
| **Dataset** | SIA, SIDIS, $pp$ | SIA | SIA | SIA, ($pp$) |
| **Hadronic species** | $\pi^\pm$, $K^\pm$, $p/\overline{p}$, $h^\pm$ | $\pi^\pm$, $K^\pm$ | $\pi^\pm$, $K^\pm$ | $\pi^\pm$, $K^\pm$, $p/\overline{p}$, ($h^\pm$) |
| **Perturbative orders** | LO, NLO | NLO | NLO | LO, NLO, NNLO |
| **Mass scheme** | ZM-VFNS | ZM-VFNS | ZM-VFNS | ZM-VFNS |

- **Many more** on the market (see *e.g.* http://lapth.cnrs.fr/ffgenerator/ and http://www2.pv.infn.it/~radici/FFdatabase/).
- More hadronic species available ($D^*$, $\Lambda$, etc.).
- HKKS set not publicly available.

# Most recent determinations

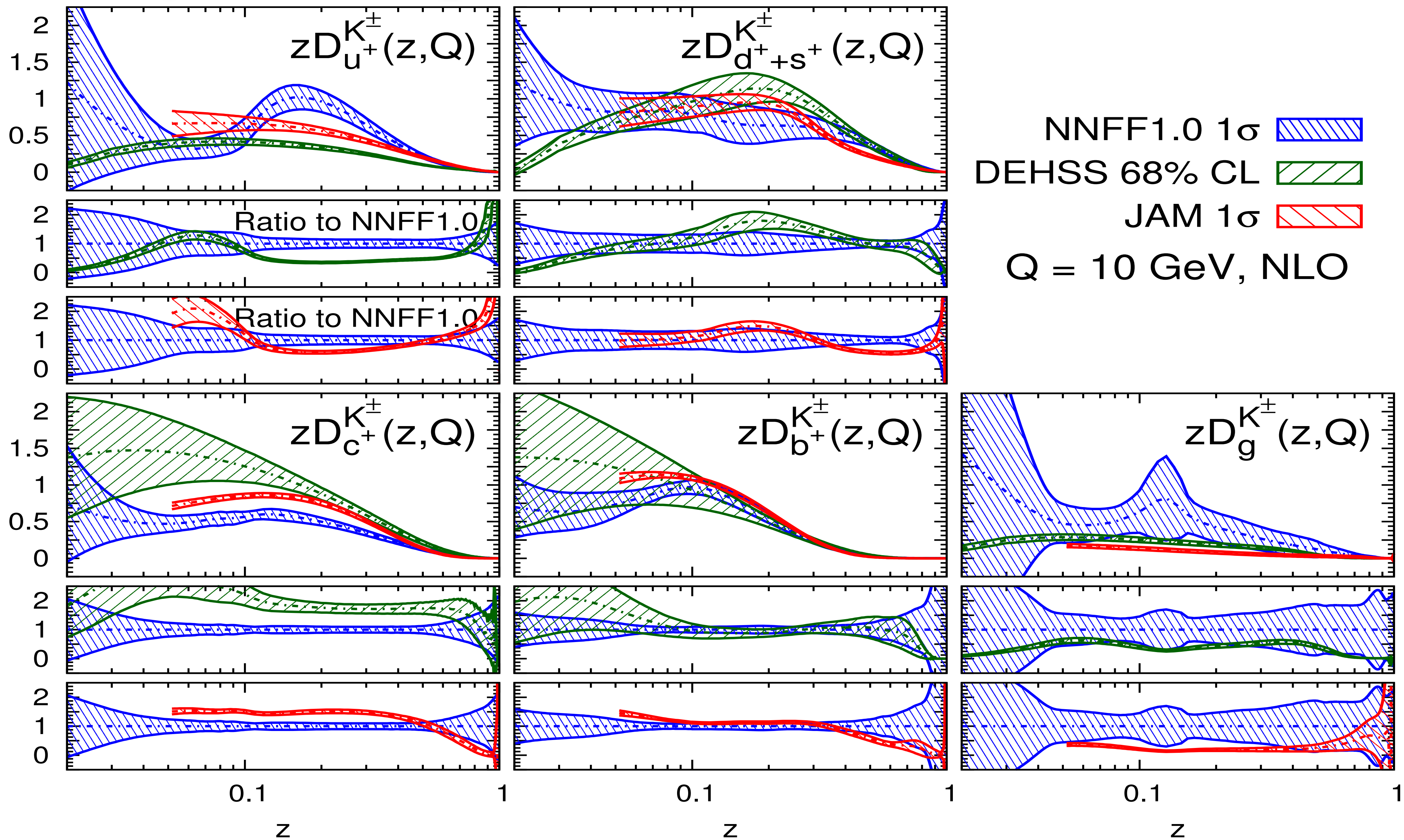
## *"Global" fits: Comparison for pions*




- Fair agreement between **NNFF1.0** (**red**) and **DEHSS** (**green**):
  - differences $u^+$ and gluon distributions at the **one-σ level**.
- More sizeable differences between **NNFF1.0** and **JAM** (**blue**):
  - the $d^+ + s^+$ and gluon distributions **well beyond one-σ**,
  - generally **smaller uncertainties** of the JAM (despite similar dataset).

# Most recent determinations

*"Global" fits: Comparison for kaons*

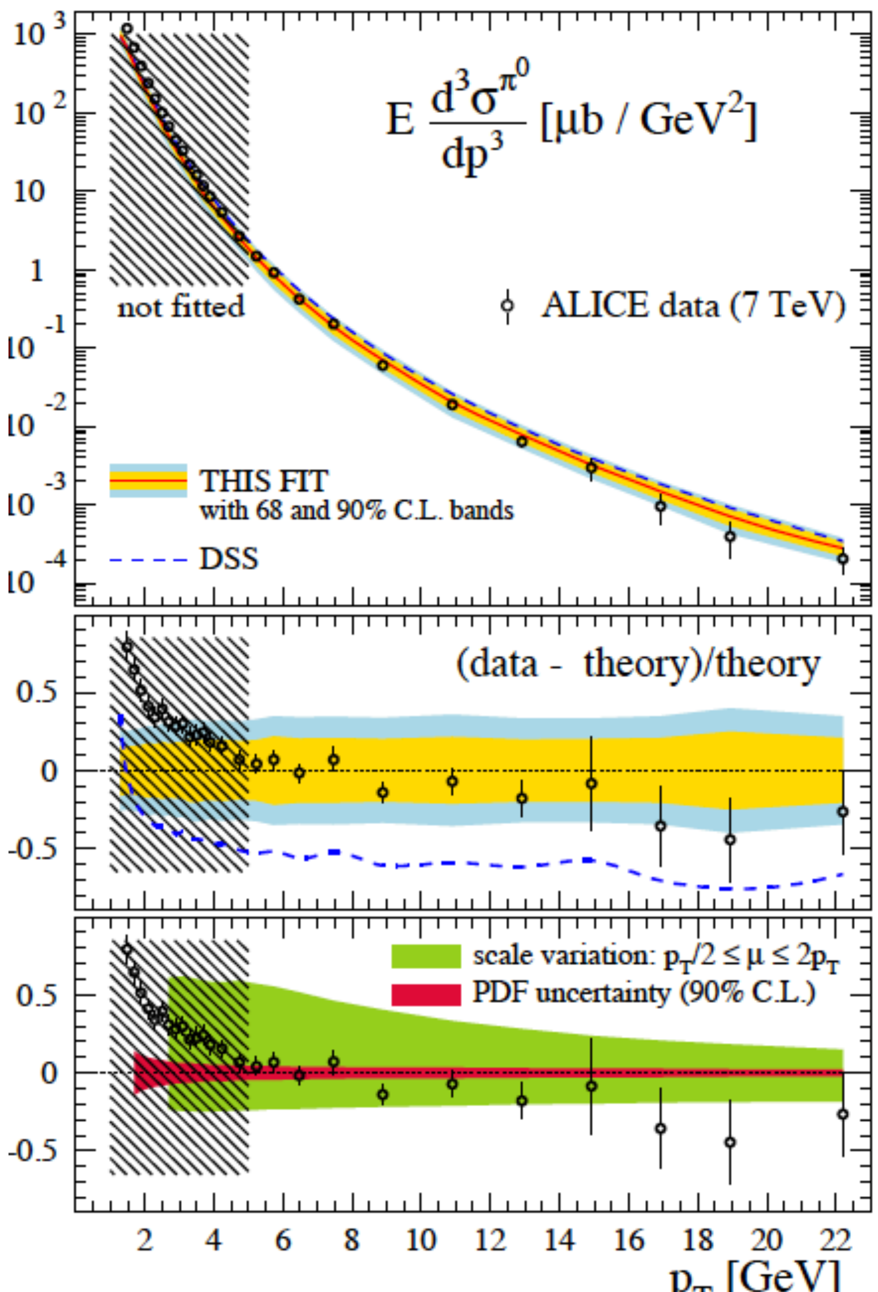



- More substantial differences all over the board as compared to pions:
  - fair agreement only for $b^+$,
  - **larger differences in the uncertainties**,
  - particularly marked for the **gluon** distributions.

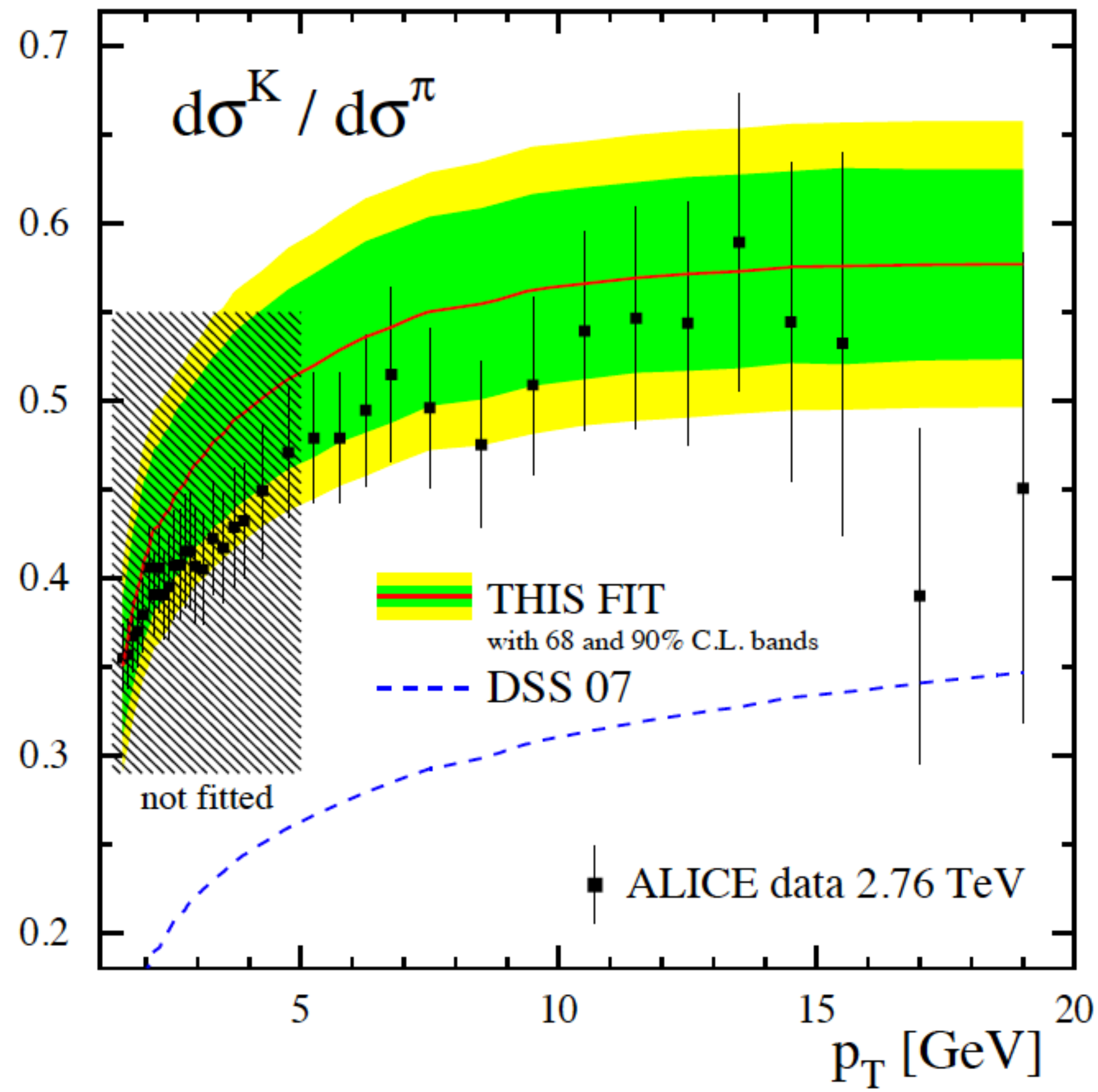
# Recent results

## *The impact of LHC data on FFs*

The DSS group included for the first time ALICE data for $\pi^0$ production at 7 TeV and K/$\pi$ ratio data at 2.76 TeV in two separate analyses for **pion** (lefts) and **kaon** (right) FFs.




de Florian *et al.* [ArXiv:1702.06353]

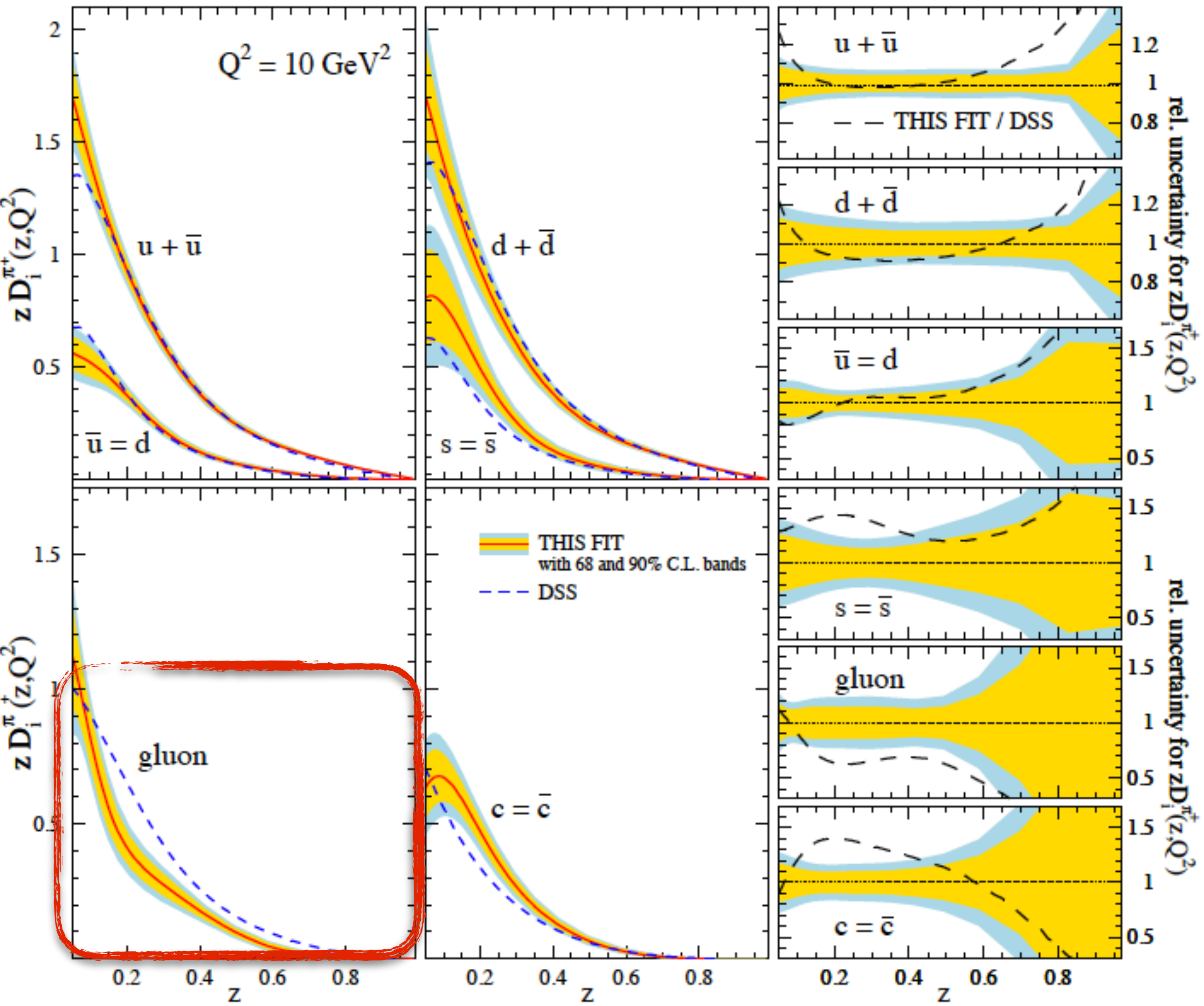



de Florian *et al.* [ArXiv:1410.6027]

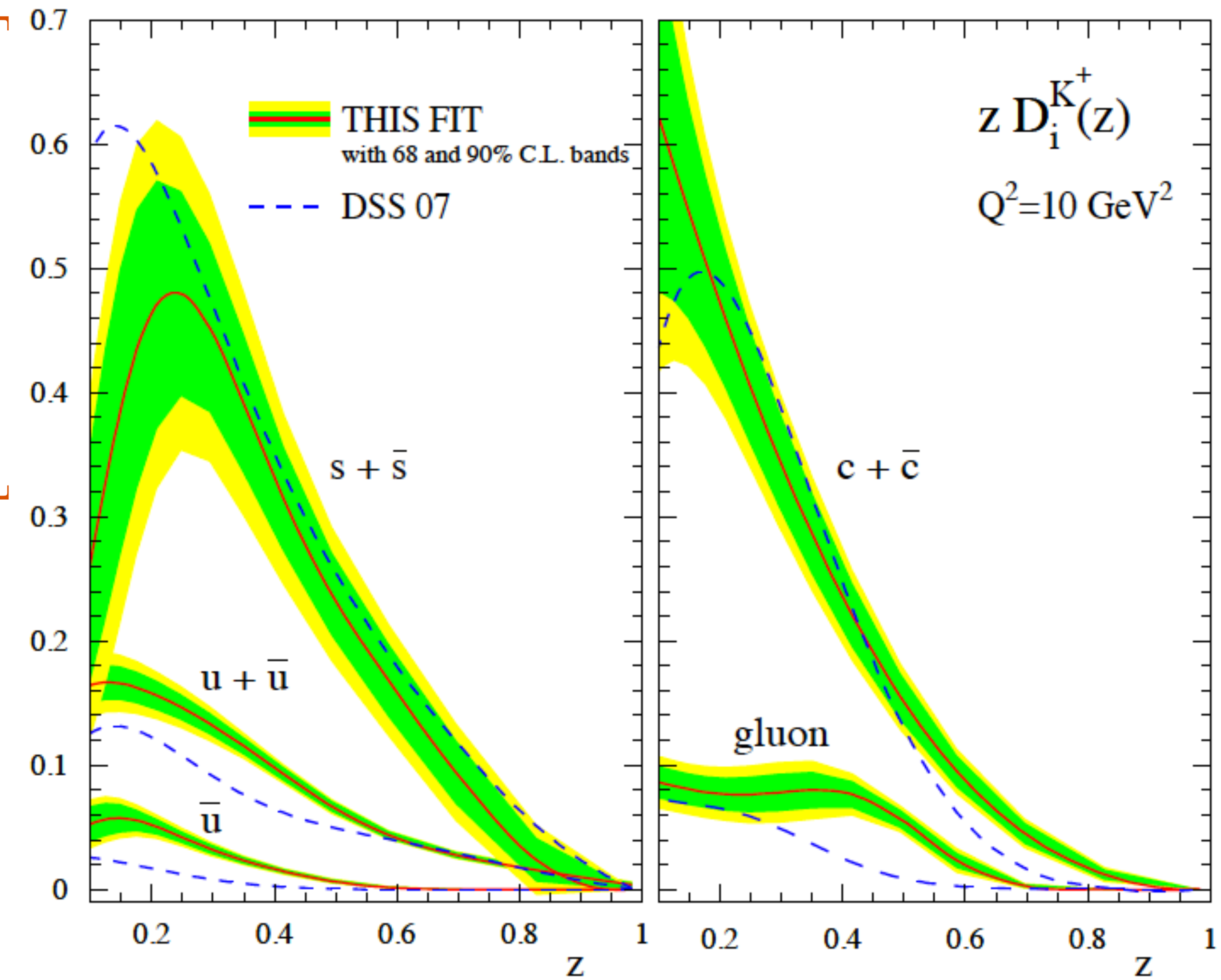
These datasets were very **poorly described** by the preceding analysis that did not include them (**DSS07**, dashed blue curve).

# Recent results

## *The impact of LHC data on FFs*




de Florian *et al.* [ArXiv:1702.06353]



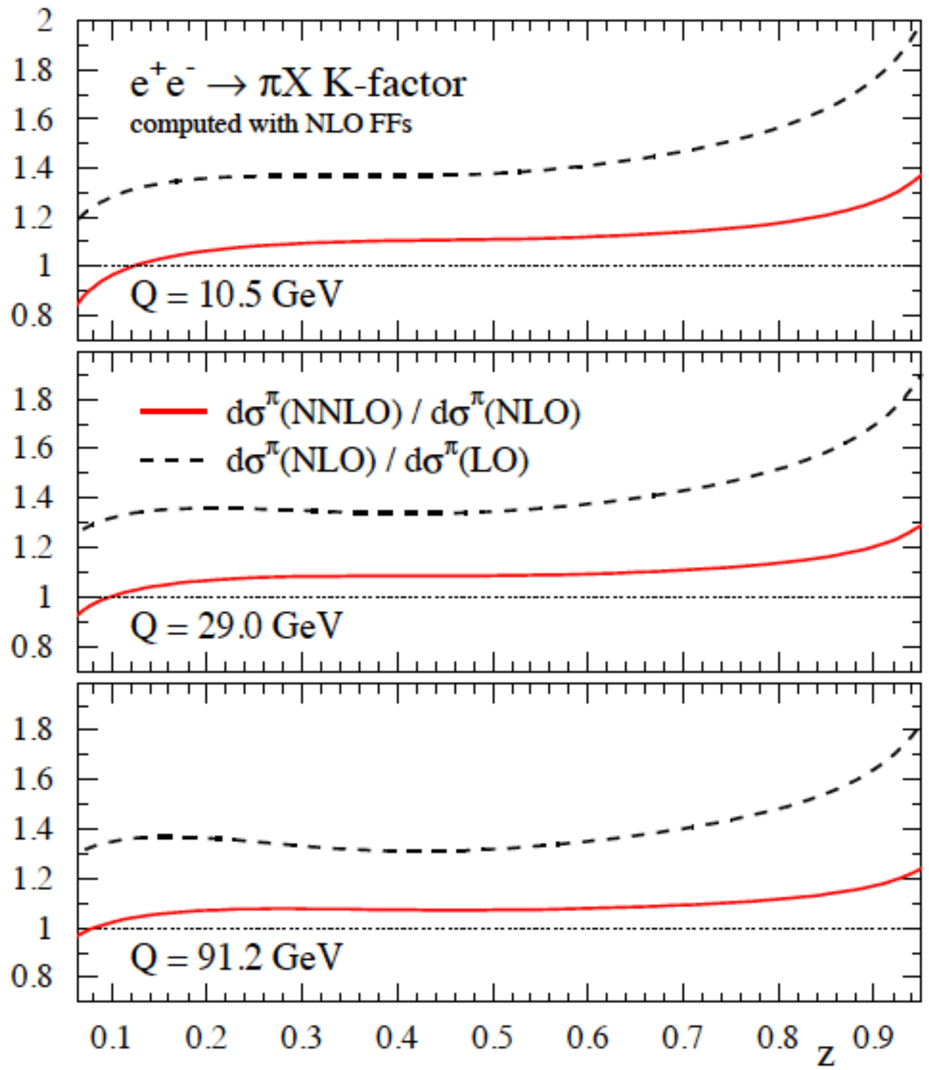

de Florian *et al.* [ArXiv:1410.6027]

- The ALICE data for **pions** (left plot) caused:
  - a strong **suppression** of the **gluon distribution**,
  - a corresponding **enhancement** of the **charm distribution**.
- Larger and more general difference for **kaons** (right plot).

# Recent results

## *The first determination of FFs at NNLO*

**NNLO** corrections to the DGLAP evolution and the hard cross sections were for the first time included in an analysis based on **SIA data** only (not available for other processes)




| experiment | data type | # data in fit | $\chi^2$ LO | NLO | NNLO |
|---|---|---|---|---|---|
| SLD [40] | incl. | 23 | 15.0 | 14.8 | 15.5 |
| | *uds* tag | 14 | 9.7 | 18.7 | 18.8 |
| | *c* tag | 14 | 10.4 | 21.0 | 20.4 |
| | *b* tag | 14 | 5.9 | 7.1 | 8.4 |
| ALEPH [41] | incl. | 17 | 19.2 | 12.8 | 12.6 |
| DELPHI [42] | incl. | 15 | 7.4 | 9.0 | 9.9 |
| | *uds* tag | 15 | 8.3 | 3.8 | 4.3 |
| | *b* tag | 15 | 8.5 | 4.5 | 4.0 |
| OPAL [43] | incl. | 13 | 8.9 | 4.9 | 4.8 |
| TPC [44] | incl. | 13 | 5.3 | 6.0 | 6.9 |
| | *uds* tag | 6 | 1.9 | 2.1 | 1.7 |
| | *c* tag | 6 | 4.0 | 4.5 | 4.1 |
| | *b* tag | 6 | 8.6 | 8.8 | 8.6 |
| BABAR [10] | incl. | 41 | 108.7 | 54.3 | 37.1 |
| BELLE [9] | incl. | 76 | 11.8 | 10.9 | 11.0 |
| NORM. SHIFTS | | | 7.4 | 6.8 | 7.1 |
| **TOTAL:** | | 288 | 241.0 | 190.0 | 175.2 |

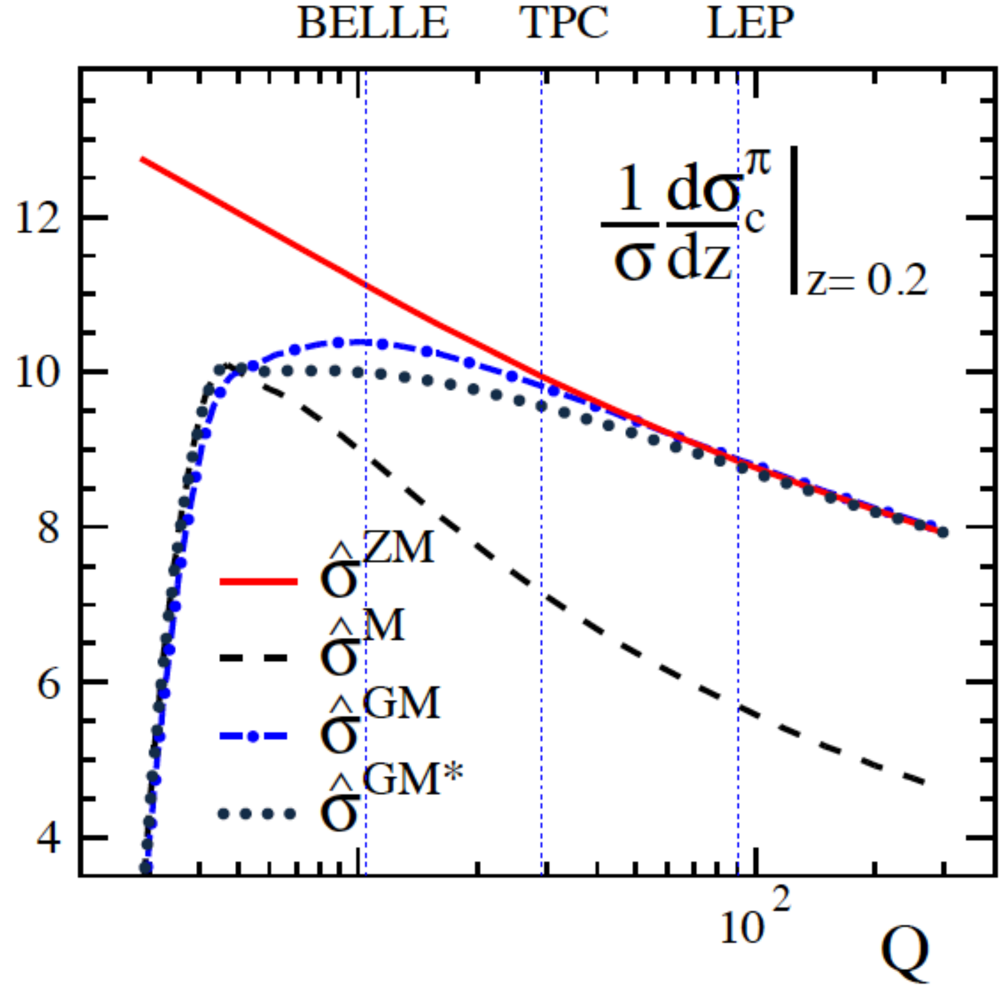
Anderle *et al.* [ArXiv:1510.05845]

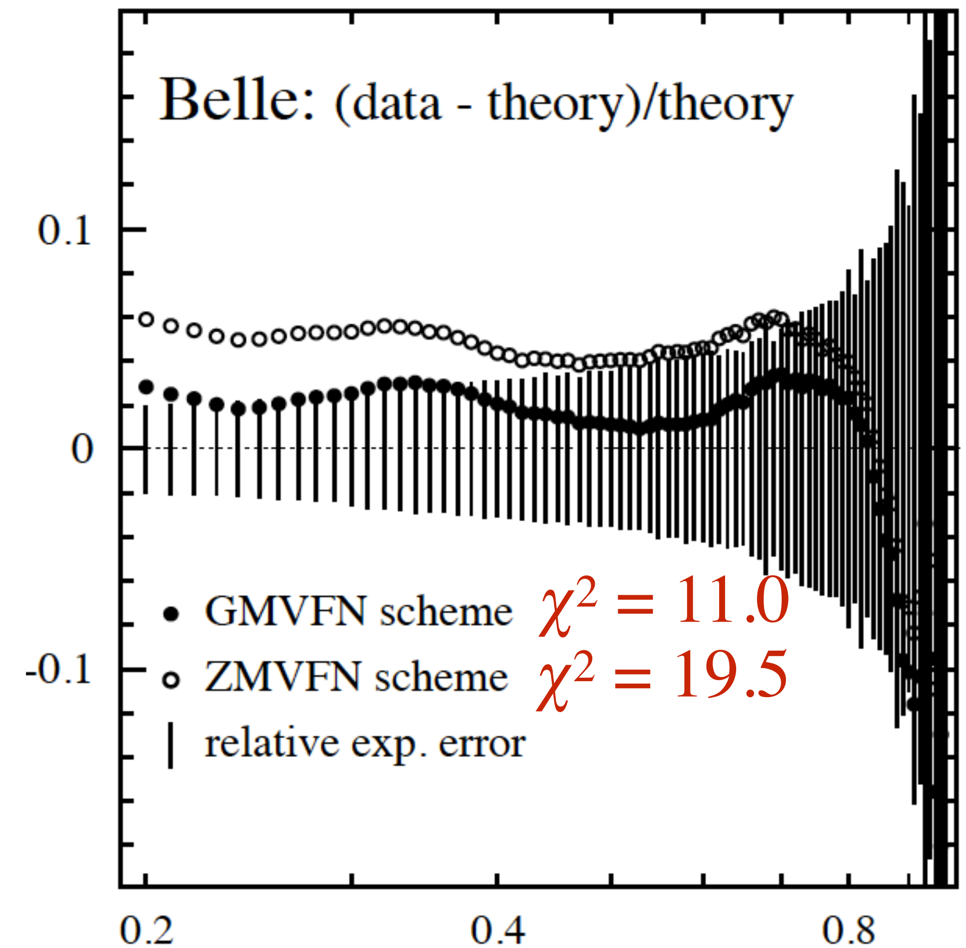
- **Large** NNLO/NLO **K-factors** (O(10%)) in the **B-factory** region ($Q \simeq 10.5$ GeV),
- Marked improvement of the global $\chi^2$ upon inclusion of higher-order corrections:
  - mostly driven by **BABAR** when moving from NLO to NNLO.

# Recent results

## *Impact of mass corrections on FFs*

**Heavy-quark mass** corrections relevant for $Q \gtrsim m_h$. The precision of the B-factories data and the kinematic coverage ($Q \simeq 2m_b$) is such that these corrections are significant.




Epele *et al.* [ArXiv:1604.08427]

- Mass corrections known to $O(\alpha_s)$ for **SIA**.
- Use of **FONLL** to smoothly interpolate between massive and massless scheme:
  - appropriate both for B-factories ($Q \simeq 10.5$ GeV) and LEP ($Q = M_Z$) data.
- Marked improvement of the $\chi^2$ of **BELLE** (and BABAR).

# Recent results

## *The impact of small-z resummation corrections*

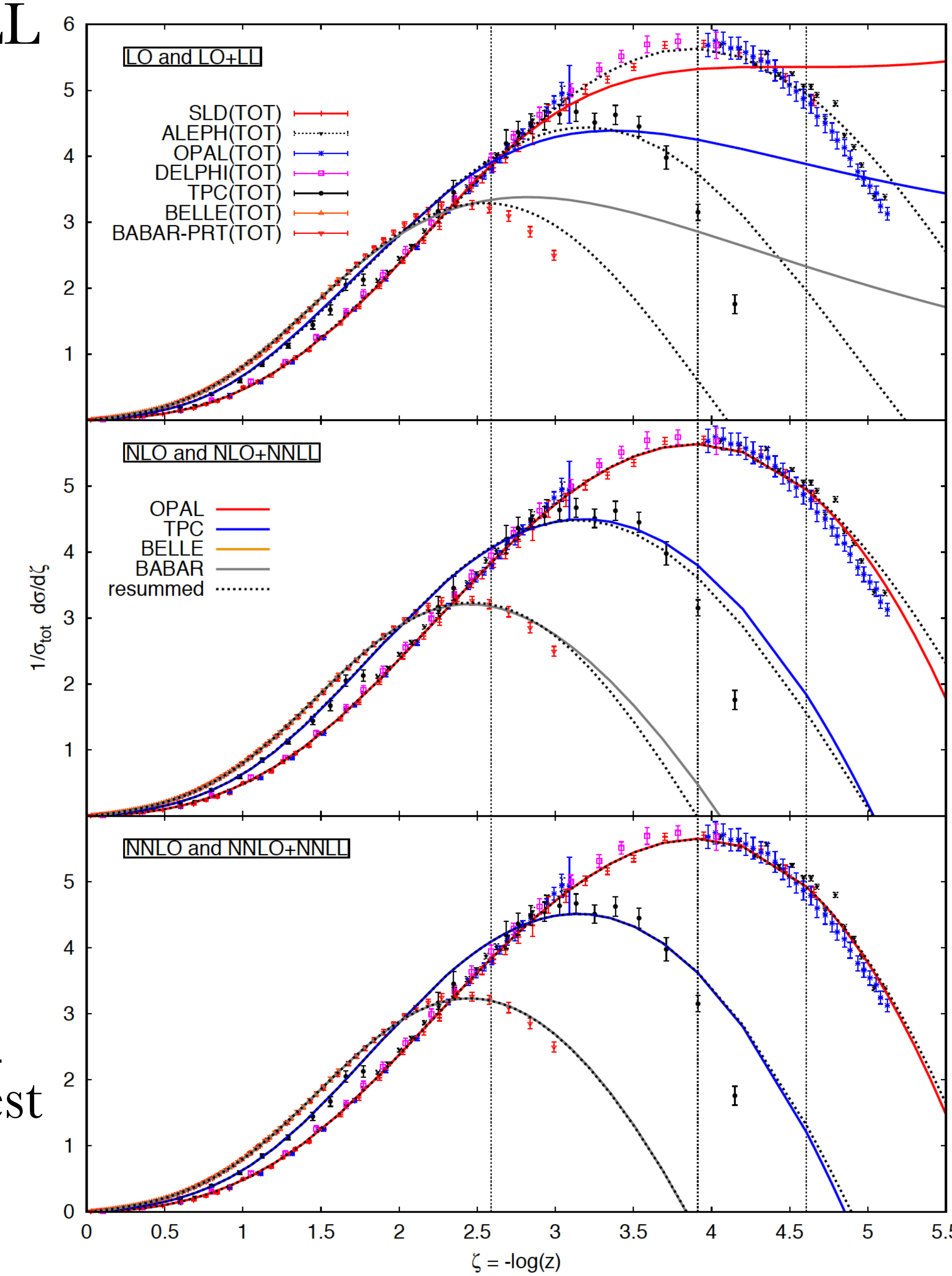
- **Small-*z* resummation** corrections up to NNLL are implemented in the **SIA** hard cross sections and the **DGLAP** splitting functions.

- **Fits** at the different orders are performed to a variety of **pion** SIA data.

| accuracy | $\chi^2$ |
|---|---|
| LO | 1260.78 |
| NLO | 354.10 |
| NNLO | 330.08 |
| LO+LL | 405.54 |
| NLO+NNLL | 352.28 |
| NNLO+NNLL | 329.96 |

- Beyond LO, resummation provides only a **small improvement** in the kinematic region of interest w.r.t. the to fixed-order (particularly at NNLO).

- Important to study the interplay with **hadron mass corrections**, effective in the same region.




Anderle *et al.* [ArXiv:1611.03371]

# Charged hadron FFs

*A brief overview*

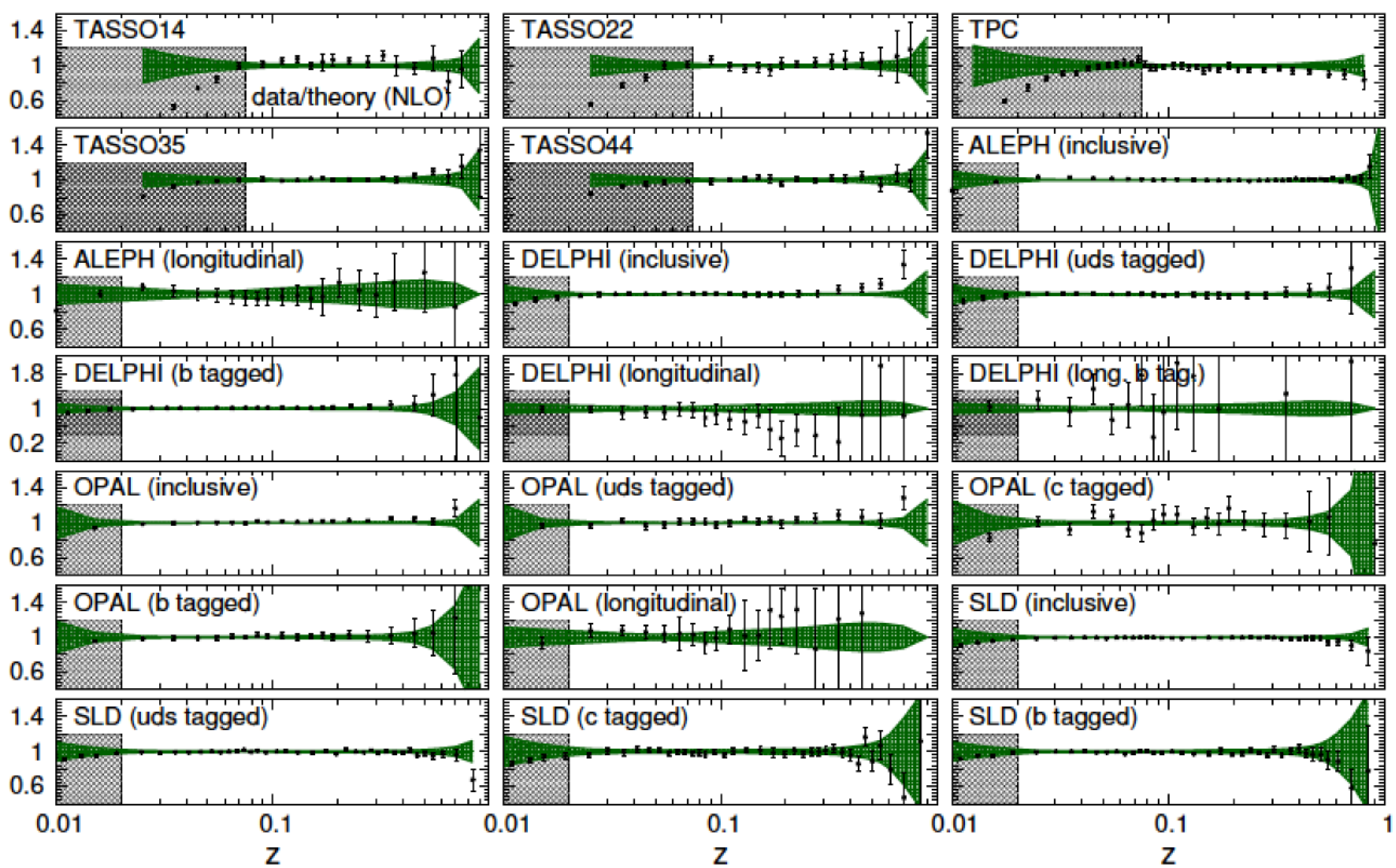
- Many experiments provide data for **charged-hadron** production:
  - this data includes, not only pions, kaons and protons, but **also heavier** (and less abundant) **charged hadrons**.
- Restricting to **SIA experiments**, data is available from:
  - TASSO, TPC, ALEPH, DELPHI, OPAL, SLD.
- Some experiments measure also the **longitudinal** cross section:
  - ALEPH, DELPHI, OPAL.
- Predictions for the longitudinal cross section start at $O(\alpha_s)$:
  - as a consequence it is **not possible to go beyond NLO** (i.e. $O(\alpha_s^2)$) yet.
  - This data provides a strong handle on the **gluon distribution**.

# Charged hadron FFs

*The NNPDF analysis*

E. Nocera [ArXiv:1709.03400]

- General good description of the entire dataset ($\chi^2 / N_{\mathrm{dat}} = 0.83$).
- Particularly good the description of the **longitudinal data**.

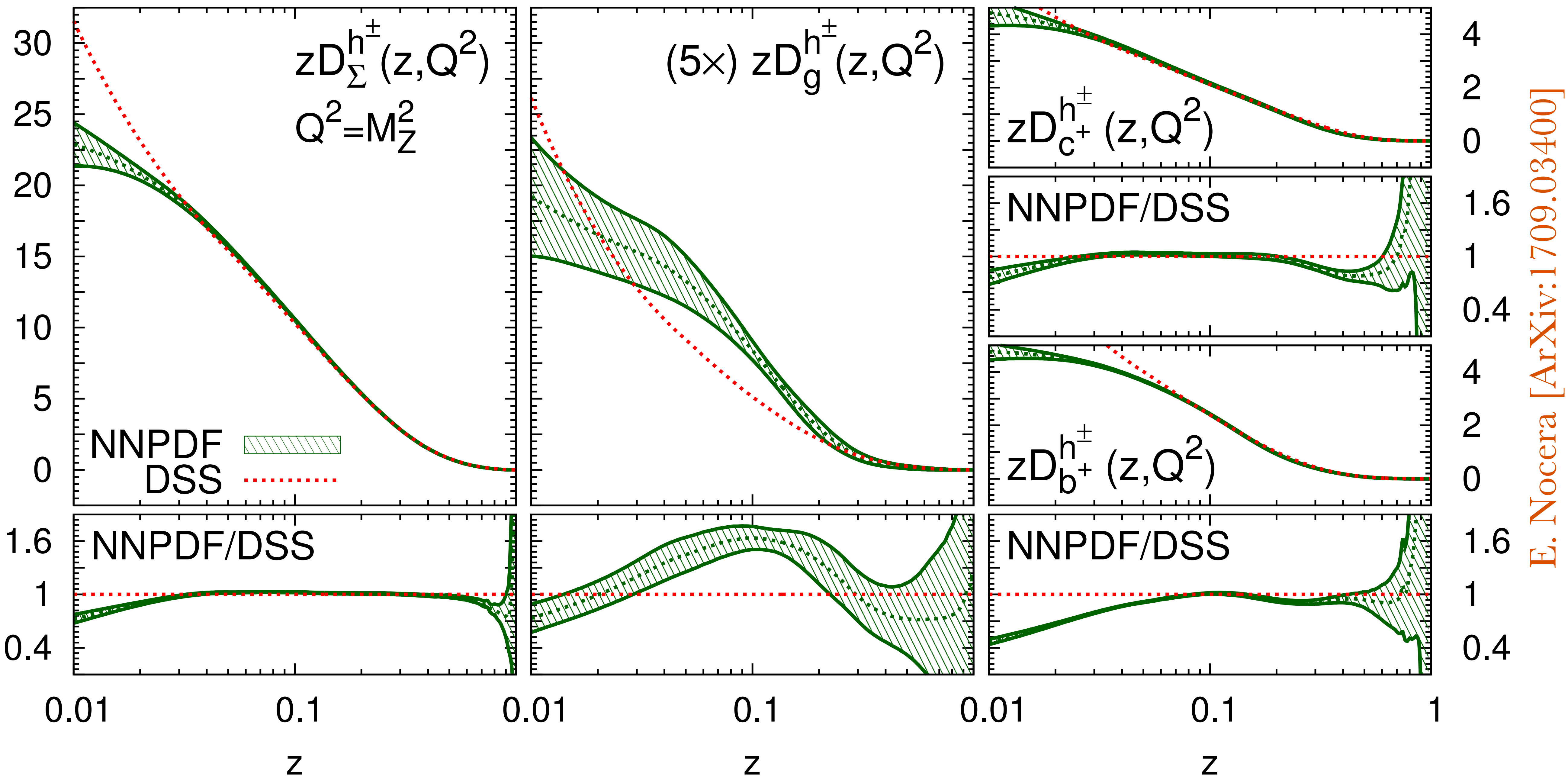



| Experiment | Reference | Observable | $\sqrt{s}$ [GeV] | $N_{\mathrm{dat}}$ | $\chi^2/N_{\mathrm{dat}}$ |
|---|---|---|---|---|---|
| TASSO14 | [5] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 14.00 | 15 (20) | 1.23 |
| TASSO22 | [5] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 22.00 | 15 (20) | 0.51 |
| TPC | [6] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 29.00 | 21 (34) | 1.65 |
| TASSO35 | [5] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 35.00 | 15 (20) | 1.14 |
| TASSO44 | [5] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 44.00 | 15 (20) | 0.68 |
| ALEPH | [7] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 91.20 | 32 (35) | 1.04 |
| | [7] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma_L^{h^\pm}}{dz}$ | 91.20 | 19 (21) | 0.36 |
| DELPHI | [8] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dp_h}$ | 91.20 | 21 (27) | 0.65 |
| | [8] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dp_h}\Big\vert_{uds}$ | 91.20 | 21 (27) | 0.17 |
| | [8] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dp_h}\Big\vert_{b}$ | 91.20 | 21 (27) | 0.82 |
| | [9] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma_L^{h^\pm}}{dz}$ | 91.20 | 20 (22) | 0.72 |
| | [9] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma_L^{h^\pm}}{dz}\Big\vert_{b}$ | 91.20 | 20 (22) | 0.44 |
| OPAL | [10] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}$ | 91.20 | 20 (22) | 2.41 |
| | [10] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{uds}$ | 91.20 | 20 (22) | 0.90 |
| | [10] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{c}$ | 91.20 | 20 (22) | 0.61 |
| | [10] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{b}$ | 91.20 | 20 (22) | 0.21 |
| | [11] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma_L^{h^\pm}}{dz}$ | 91.20 | 20 (22) | 0.31 |
| SLD | [12] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dp_h}$ | 91.28 | 34 (40) | 0.75 |
| | [12] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{uds}$ | 91.28 | 34 (40) | 1.03 |
| | [12] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{c}$ | 91.28 | 34 (40) | 0.62 |
| | [12] | $\frac{1}{\sigma_{\mathrm{tot}}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{b}$ | 91.28 | 34 (40) | 0.97 |
| Total dataset | | | | 471 (527) | 0.83 |

# Charged hadron FFs

*The NNPDF analysis*

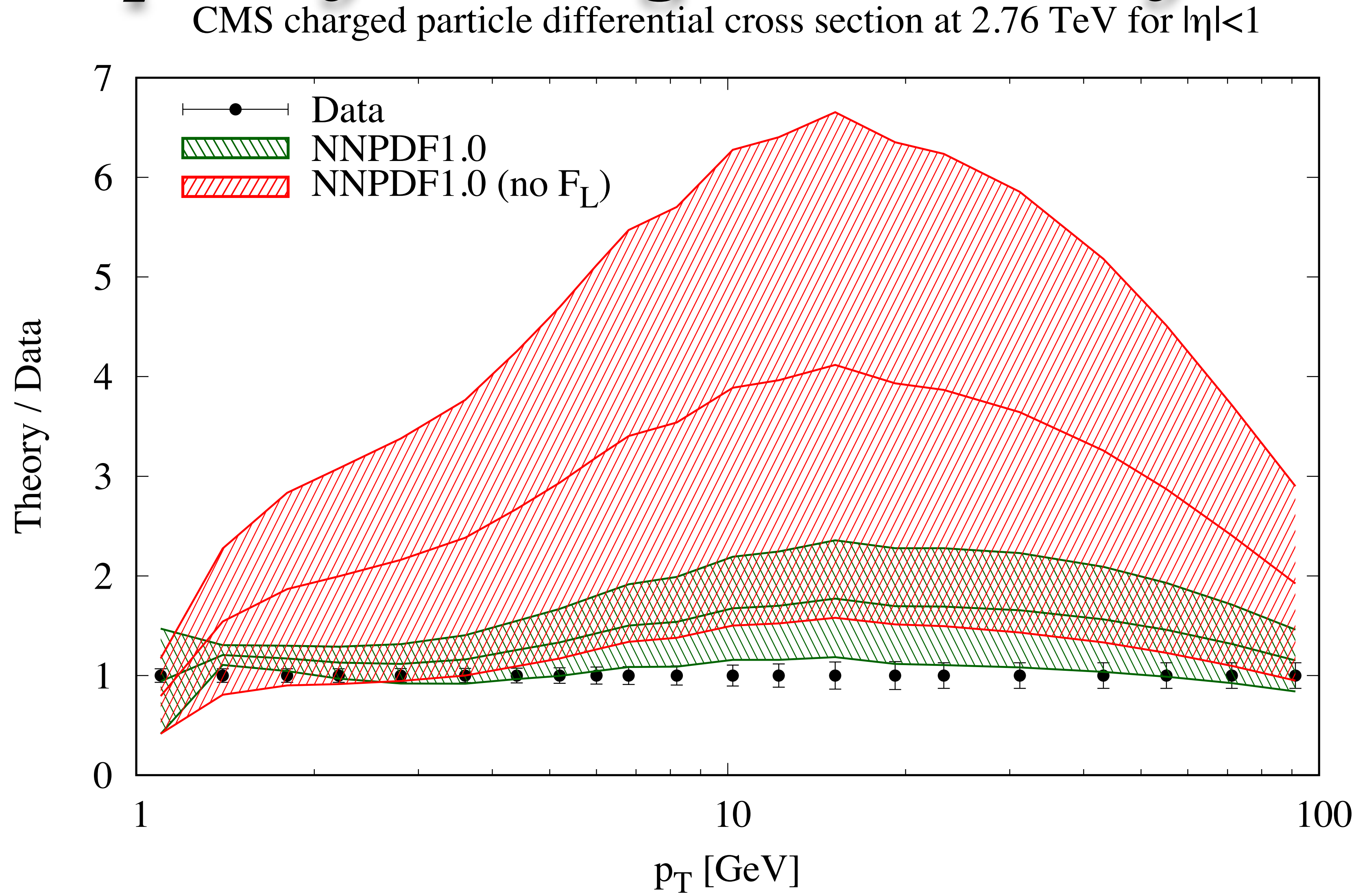
- Charged hadron FFs at the Z-boson mass scale:




- Significant differences w.r.t. DSS, particularly for the gluon.

# Charged hadron FFs

## *Aside: the impact of the longitudinal data for the LHC*




- Strong **sensitivity to the gluon distribution**.
- Very significant impact of the **longitudinal data**:
  - dramatic reduction of the **uncertainty**,
  - **better agreement** with CMS data.
- **LHC** and **Tevatron** data expected to have a big impact on FFs.

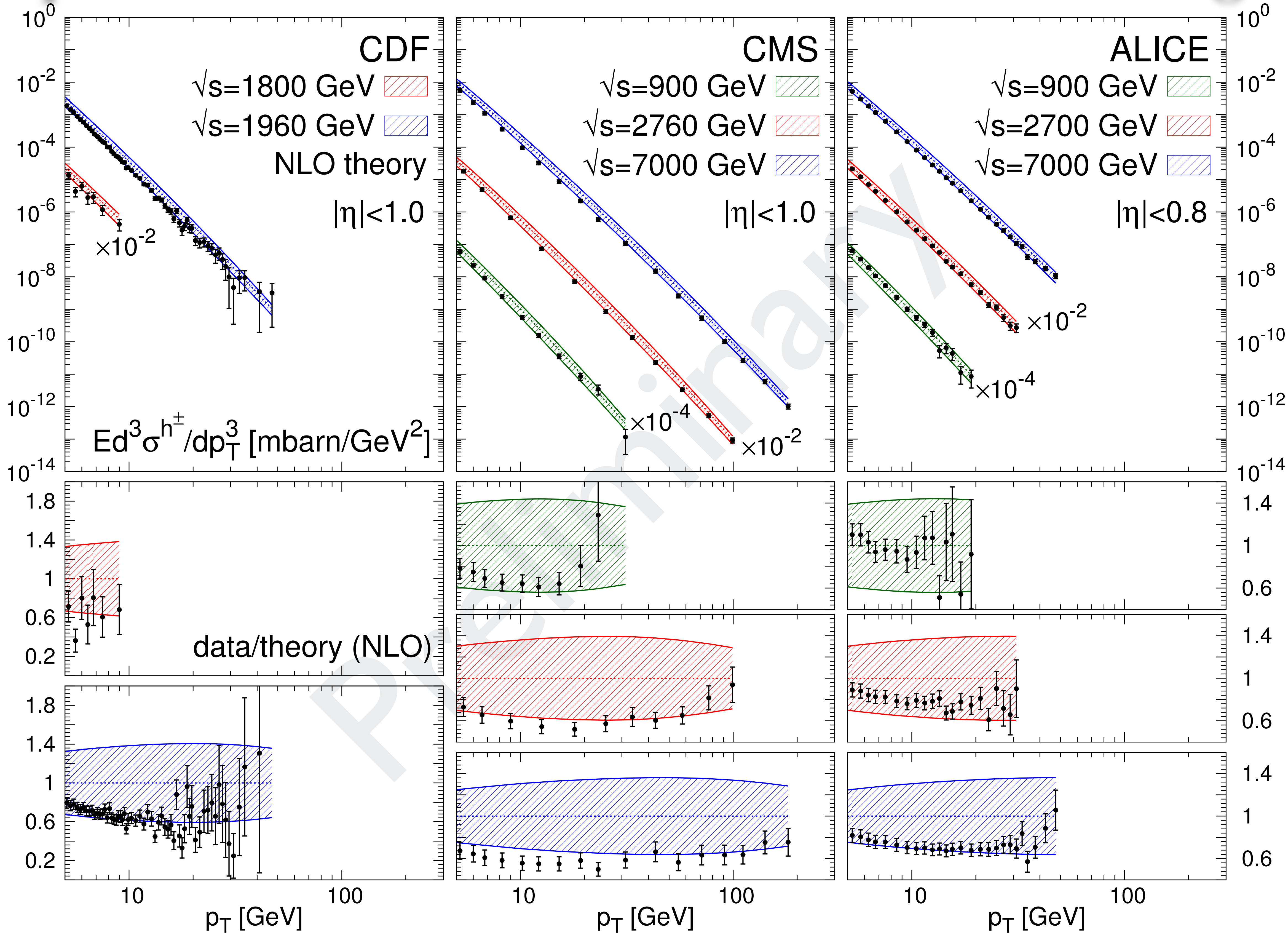
# Charged hadron FFs

*Impact hadroproduction in pp collision data (preliminary)*

- **CDF** at the Tevatron, and **CMS** and **ALICE** at the LHC released charged-hadron $p_T$ spectra at different c.o.m. energies:
  - CMS and ALICE at $\sqrt{S}$ = 900, 2760, and 7000 GeV,
  - CDF at $\sqrt{S}$ = 630, 1800, and 1960 GeV.
  - Sensitivity to the **charged-hadron FFs**, particularly to the **gluon**,
- Hard cross sections currently know to **NLO** (i.e. $O(\alpha_s^3)$).
  - **large scale variations** at low $p_T$. Consider only data with $p_T > 5$ GeV.
  - No CDF data at 630 GeV survives.
- Include CMS, ALICE, and CDF data in the NNFF1.0 analysis of charged-hadron FFs by means of **Bayesian reweighting**:
  - use NNPDF31_nlo_as_0118 for the PDFs,
  - only **preliminary** results for now.

# Charged hadron FFs

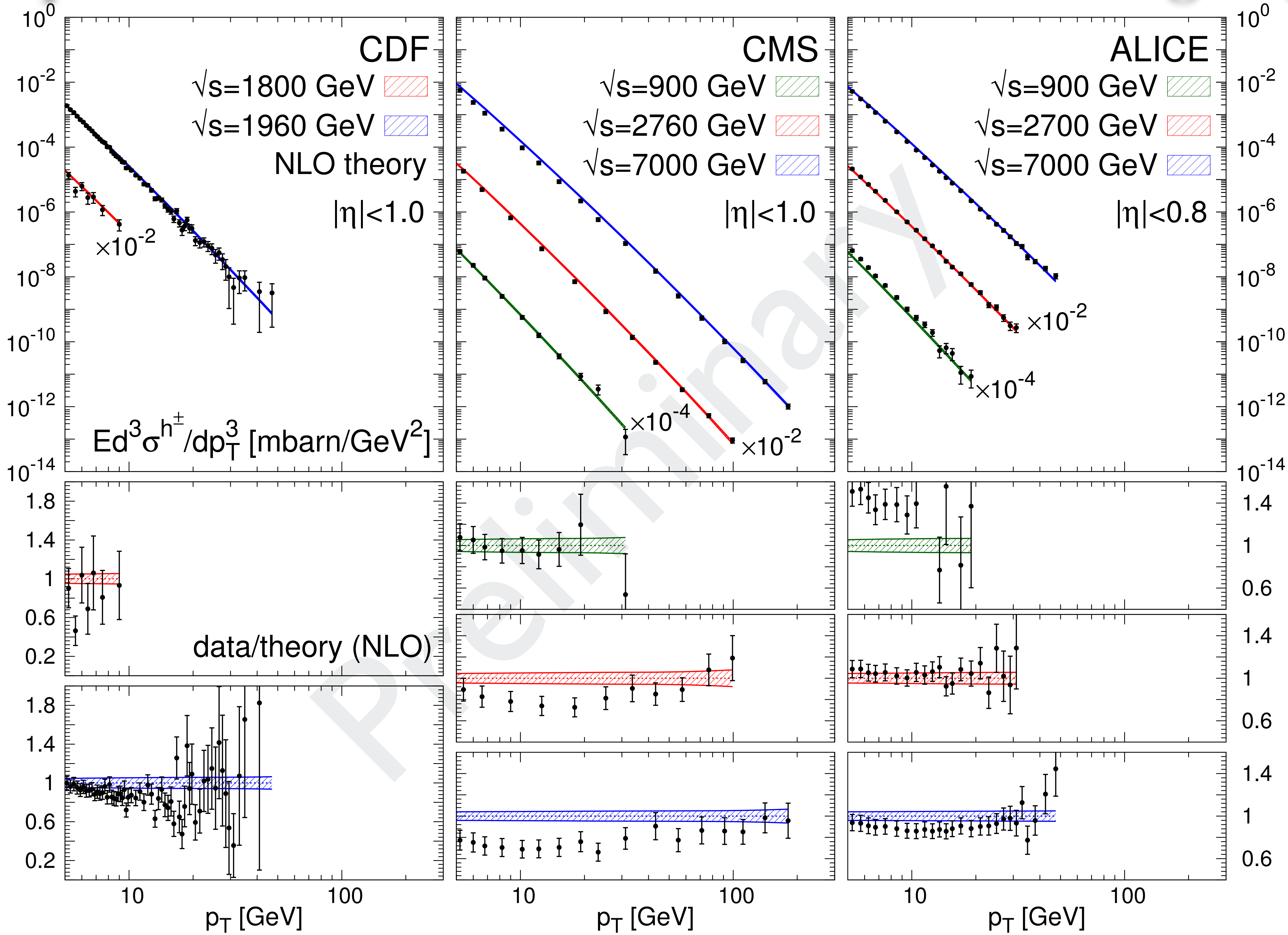
*Comparison to LHC/Tevatron data BEFORE reweighting*




[Thanks to E. Nocera]

# Charged hadron FFs

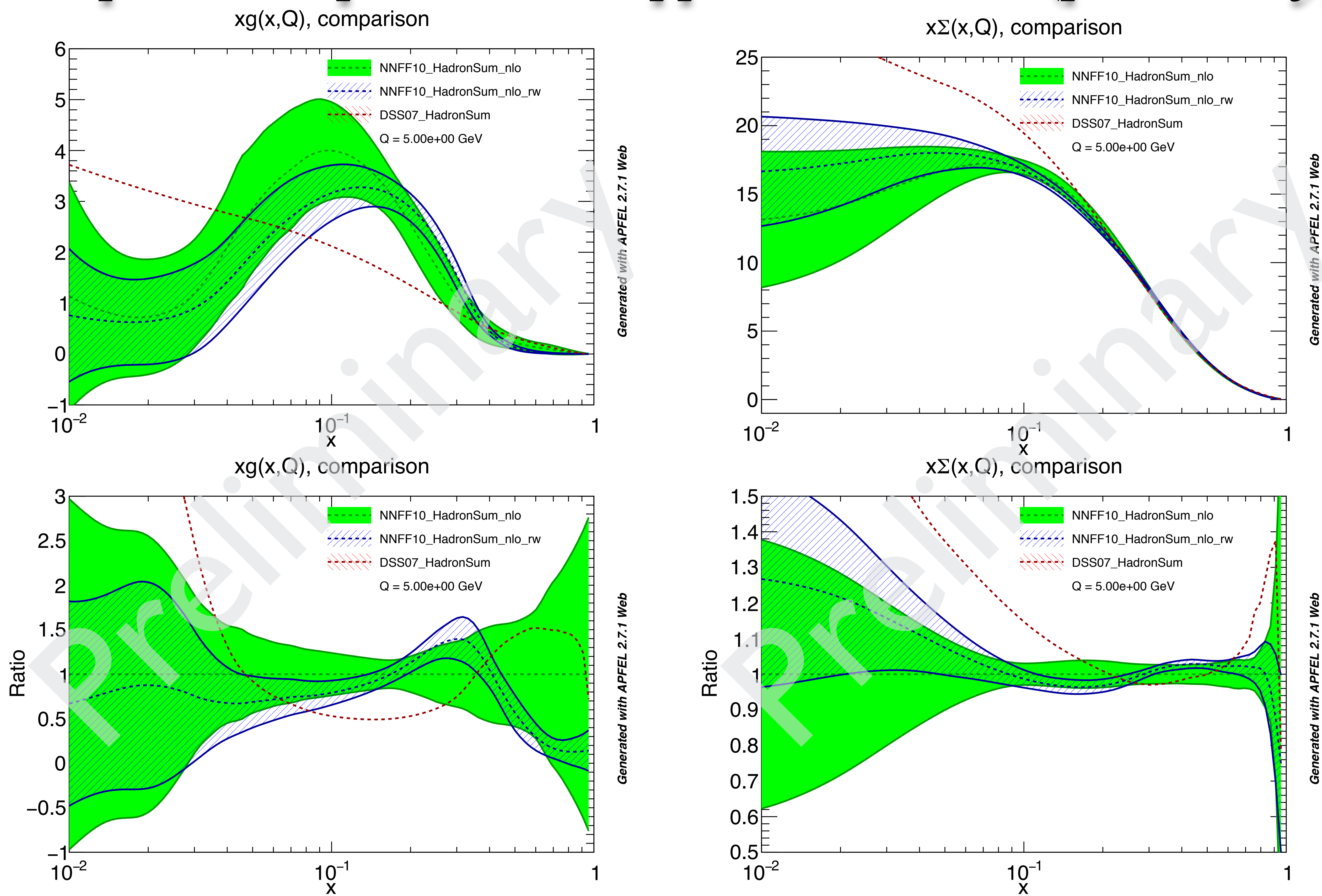
*Comparison to LHC/Tevatron data AFTER reweighting*




[Thanks to E. Nocera]

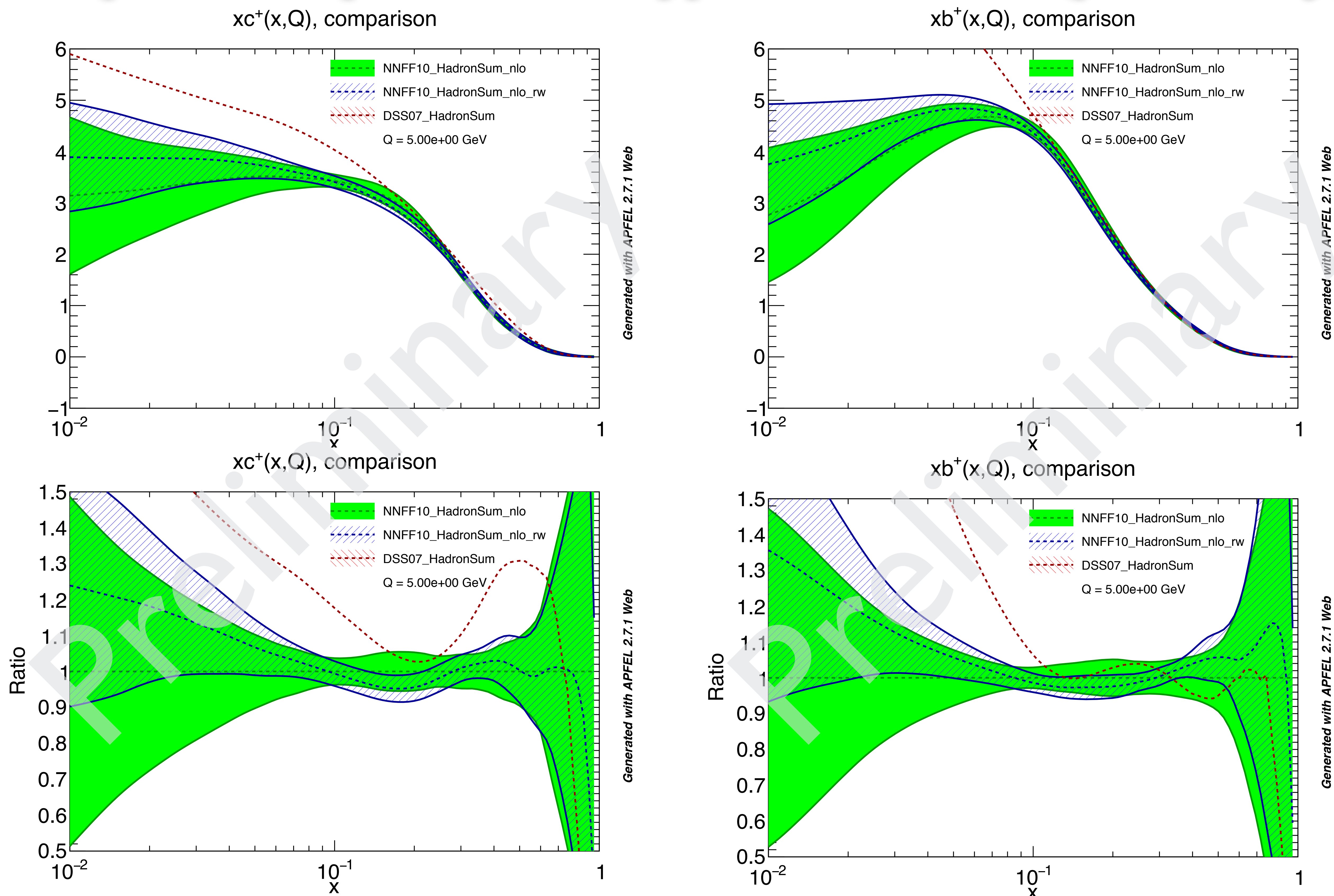
# Charged hadron FFs

## *Impact hadroproduction in pp collision data (preliminary)*



# Charged hadron FFs

*Impact hadroproduction in pp collision data (preliminary)*



# Summary

- **Well-established** procedure in the context of **collinear factorisation** to extract non-perturbative quantities from fits to **experimental data**:
  - **PDFs** have driven many of the technical developments in this context but in the last years **FFs** are quickly catching up.
  - An impressive amount of **new results** have been produces recently in the field of FFs (only a few were quickly mentioned).
- The **intrinsic complexity** of the procedure leaves room for different choices that make up most of the difference between the determinations:
  - profitable **exchange** between collaborations.
- On the **experimental side**, a wealth of new precise data from:
  - BELLE, and BABAR (SIA),
  - HERMES and COMPASS (SIDIS),
  - Tevatron, LHC, RHIC (*pp*).
- Many developments on the **theory side**:
  - higher-order corrections,
  - resummation corrections,
  - heavy-quark mass corrections.

# Items for discussion

- **PDFs** and **FFs** are often simultaneously involved in the computation of observables (*e.g.* SIDIS and hadroprodution in *pp* collisions).

- A **simultaneous** determination of PDFs and FFs is probably the way to go for a full exploitation of the experimental data.

- So far, most of the determinations of FFs including SIDIS and *pp* data relied on a preexisting set of PDFs (*e.g.* MMHT for DSS):

  - a first attempt of a simultaneous determination of (polarised) PDFs and FFs was recently carried out (J. J. Ethier *et al.* [ArXiv:1705.05889]) with encouraging results.

- In my **strictly personal opinion** a sound simultaneous determination of (polarised and/or unpolarised) PDFs and FFs from a **global** dataset requires pushing the methodological frontiers beyond the current status:

  - **fast** and **reliable** (i.e. benchmarked) computational tools are needed.

    - The widespread **Mellin** method is fast but typically requires simple functional forms to be effectively performing. Is that ok? Probably not for unpolarised PDFs.

    - The ***x*-space** formulation can be made as fast as the Mellin one with no restriction on the functional forms, but requires work on the development of suitable tools (*e.g.* APPLgrid and FastNLO). *pp* is particularly problematic given the presence of *three* convolutions.

# Items for discussion

- The choice of a **suitable parameterisation** is crucial:
  - a potentially large number of parameters. Can simple minimisation algorithms cope with a very complex parameter space? Maybe not. Quoting the MINUIT manual (http://inspirehep.net/record/1258345/files/mnusersguide.pdf):

    > according to the actual needs and "on demand". There is no protection against an upper limit on the number of parameters, however the "technological" limitations of MINUIT can be seen around a maximum of 15 free parameters at a time.

  - There appears to be an *intrinsic limitation* to the complexity of the parameter space that a "standard" minimisation algorithm can deal with.
  - **Machine learning** techniques can come to rescue. NNPDF is currently using (and exploring) some of them.
  - A suitable parameterisation is also vital for a reliable **propagation of the experimental uncertainties**. Not only in the extrapolation regions, but also in the data region.
- The full **exploitation of the experimental information** is very relevant:
  - modern experiments release data along with the full **correlation breakdown** (in the form of nuisance parameters or covariance matrix),
  - when uncorrelated uncertainties are at the percent level, **correlations** might become the dominant in the definition of the figure of merit ($\chi^2$).
  - Computation of the $\chi^2$ might become expensive. SVD techniques can help.

# Backup slides

# Momentum sum rule

$$M_i(Q) = \sum_{h=\pi^\pm, K^\pm, p/\bar{p}} \int_{z_{\min}}^{1} dz\, z D_i^h(z,Q) < 1$$

$$\begin{aligned} M_g(Q = 5\text{ GeV}) &= 0.82 \pm 0.18 \\ M_g(Q = 10\text{ GeV}) &= 0.79 \pm 0.16 \\ M_g(Q = M_Z) &= 0.70 \pm 0.12 \end{aligned}$$

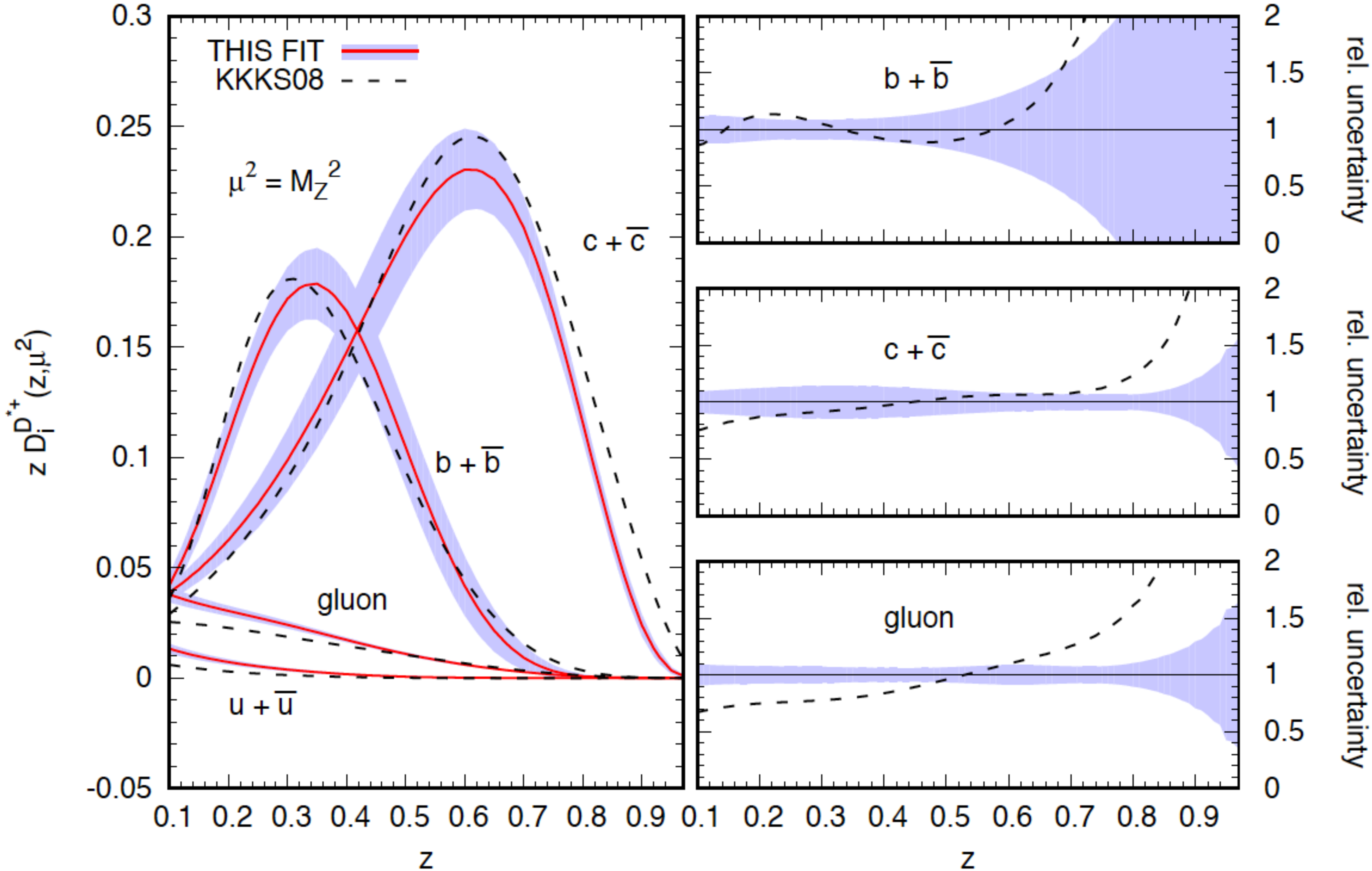
$$M_i(Q) = \sum_{h=\pi^\pm, K^\pm, p/\bar{p}} \int_{z_{\min}}^{1} dz\, z D_i^h(z,Q) < N$$

$N = 2$ for $i = u^+, c^+, b^+$ and $N = 4$ for $i = d^+ + s^+$

$$\begin{aligned} M_{u^+}(Q = 5\text{ GeV}) &= 1.41 \pm 0.13 \\ M_{d^+ + s^+}(Q = 5\text{ GeV}) &= 2.12 \pm 0.25 \\ M_{c^+}(Q = 5\text{ GeV}) &= 1.04 \pm 0.06 \\ M_{b^+}(Q = 5\text{ GeV}) &= 1.01 \pm 0.06 \end{aligned}$$

# Recent results

*A determination of the FFs of D**




Anderle *et al.* [ArXiv:1706.09857]

# Settings

- **Physical parameters**:

$$\alpha_s(M_Z) = 0.118, \quad \alpha_{\rm em}(M_Z) = 1/127, \quad m_c = 1.51 \text{ GeV}, \quad m_b = 4.92 \text{ GeV}$$

- **Parametrisation scale**:

$$Q_0 = 5 \text{ GeV} \ (> m_c, m_b)$$

  - substantial heavy-quark intrinsic component,
  - heavy-quark FFs parametrised on the same footing as the light FFs.

- **5 independent FFs** for each hadronic species $h$:

$$\left\{ D^h_{u^+}, D^h_{s^+ + d^+}, D^h_{c^+}, D^h_{b^+}, D^h_g \right\}$$

  - **inclusive SIA data** only constrains three FF combinations,
  - heavy-quark FFs constrained directly by **tagged SIA data**.

- Each FF is parametrised by a **Neural Net** (architecture 2-5-3-1).

- **Kinematic cuts**:

$$z_{\rm min} \leq z \leq z_{\rm max}, \quad z_{\rm min} = \begin{cases} 0.02 & \text{for } \sqrt{s} = M_Z \\ 0.075 & \text{otherwise} \end{cases}, \quad z_{\rm max} = 0.9$$

# Settings

- **Physical parameters**:

$$\alpha_s(M_Z) = 0.118, \quad \alpha_{\rm em}(M_Z) = 1/127, \quad m_c = 1.51 \text{ GeV}, \quad m_b = 4.92 \text{ GeV}$$

- **Parametrisation scale**:

$$Q_0 = 5 \text{ GeV } (> m_c, m_b)$$

  - substantial heavy-quark intrinsic component,
  - heavy-quark FFs parametrised on the same footing as the light FFs.

- **5 independent FFs** for each hadronic species $h$:

$$\left\{ D^h_{u^+}, D^h_{s^+ + d^+}, D^h_{c^+}, D^h_{b^+}, D^h_g \right\}$$

  - **inclusive SIA data** only constrains three FF combinations,
  - heavy-quark FFs constrained directly by **tagged SIA data**.

- Each FF is parametrised by a **Neural Net** (architecture 2-5-3-1).

- **Kinematic cuts**: contributions $\propto M_h/ sz^2$ and $\ln(z)$

$$z_{\rm min} \leq z \leq z_{\rm max}, \quad \boxed{z_{\rm min} = \begin{cases} 0.02 & \text{for } \sqrt{s} = M_Z \\ 0.075 & \text{otherwise} \end{cases}} , \quad z_{\rm max} = 0.9$$

# Settings

- **Physical parameters**:

$$\alpha_s(M_Z) = 0.118, \quad \alpha_{\rm em}(M_Z) = 1/127, \quad m_c = 1.51 \text{ GeV}, \quad m_b = 4.92 \text{ GeV}$$

- **Parametrisation scale**:

$$Q_0 = 5 \text{ GeV} \ (> m_c, m_b)$$

  - substantial heavy-quark intrinsic component,
  - heavy-quark FFs parametrised on the same footing as the light FFs.

- **5 independent FFs** for each hadronic species $h$:

$$\left\{ D^h_{u^+}, D^h_{s^+ + d^+}, D^h_{c^+}, D^h_{b^+}, D^h_g \right\}$$

  - **inclusive SIA data** only constrains three FF combinations,
  - heavy-quark FFs constrained directly by **tagged SIA data**.

- Each FF is parametrised by a **Neural Net** (architecture 2-5-3-1).

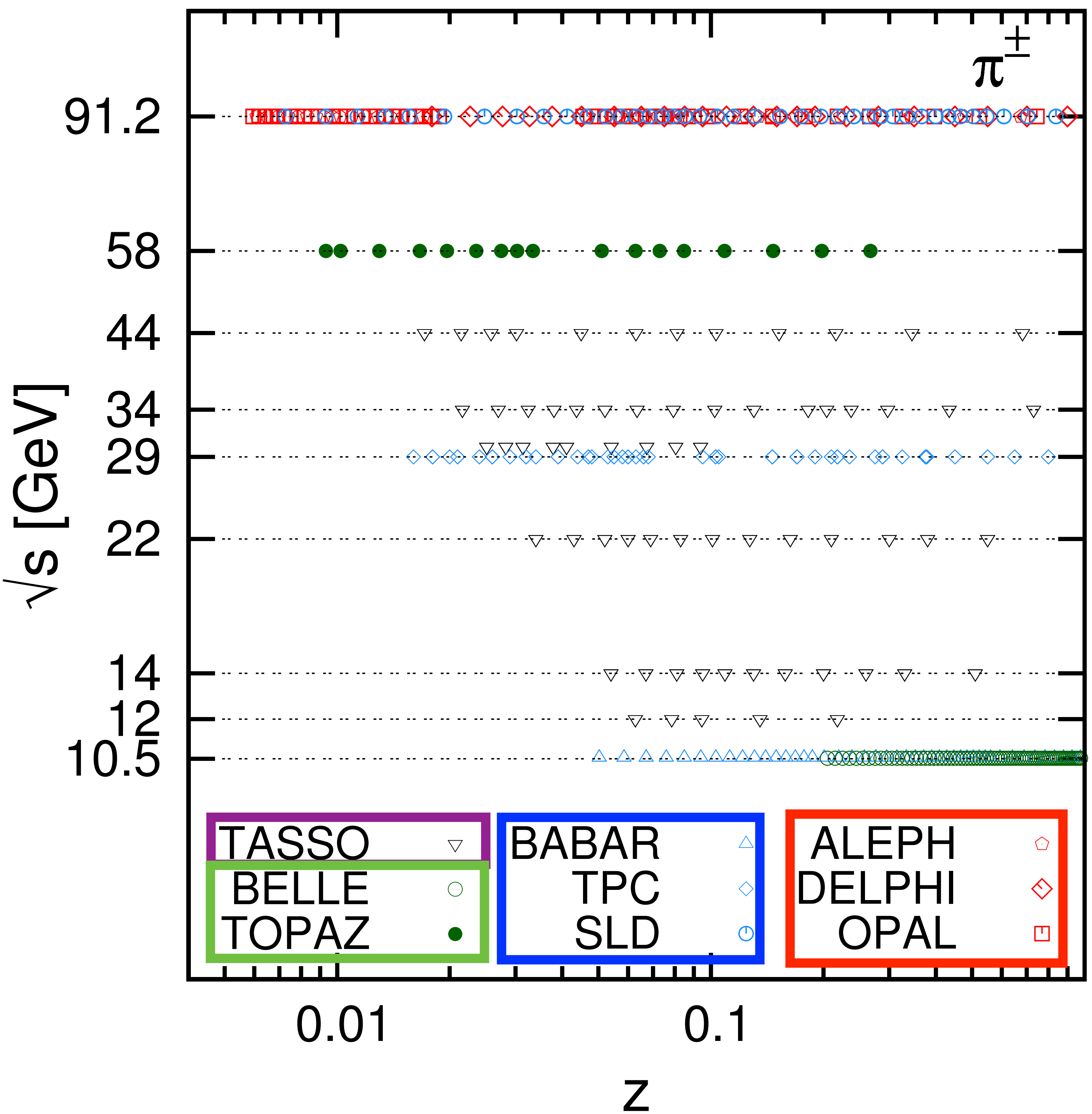
- **Kinematic cuts**:

contributions $\propto \ln(1 - z)$

$$z_{\rm min} \le z \le z_{\rm max}, \quad z_{\rm min} = \begin{cases} 0.02 & \text{for } \sqrt{s} = M_Z \\ 0.075 & \text{otherwise} \end{cases}, \quad z_{\rm max} = 0.9$$

# Dataset

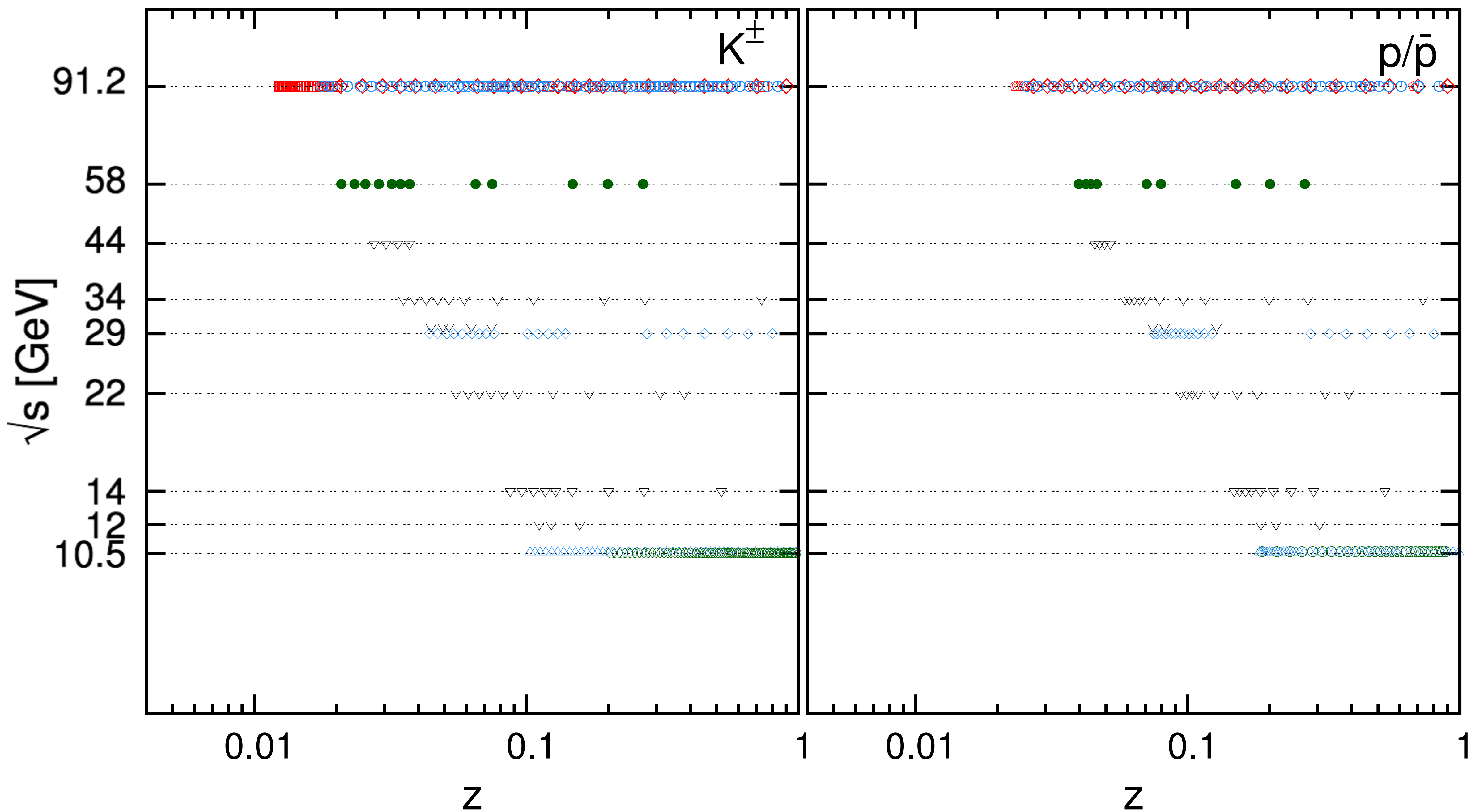
- **Only SIA cross sections** (normalised and absolute) included.




- PETRA (DESY)
  - Phys. Lett. B94 (1980) 444
  - Z.Phys. C17 (1983) 5-15
  - Z.Phys. C42 (1989) 189
- KEK
  - Phys. Rev. Lett. 111 (2013) 062002
  - Phys. Rev. D92 (2015), no. 9 092007
  - Phys. Lett. B345 (1995) 335-342
- SLAC
  - Phys. Rev. D88 (2013) 032011
  - Phys. Rev. Lett. 61 (1988) 1263
  - Phys. Rev. D69 (2004) 072003
  - X.Q. Lu, Ph.D. thesis, Johns Hopkins U., 1986
- LEP (CERN)
  - Z. Phys. C66 (1995) 355-366
  - Eur. Phys. J. C5 (1998) 585-620
  - Z. Phys. C63 (1994) 181-196.
- 428 data points (after cuts)

# Dataset

- **Only SIA cross sections** (normalised and absolute) included.
- We have fitted FFs also to $K^{\pm}$ and $p/\overline{p}$ data.




385 data points
(after cuts)

360 data points
(after cuts)

# Fit quality

- Fit quality **increasingly better** going from LO to NNLO:
  - substantial from LO to NLO, more moderate from NLO to NNLO.
  - **NNLO** corrections are anyway **beneficial** (particularly for pions).

| Exp. | $\chi^2/N_{\rm dat}$ $(h=\pi^\pm)$ LO | NLO | NNLO | $\chi^2/N_{\rm dat}$ $(h=K^\pm)$ LO | NLO | NNLO | $\chi^2/N_{\rm dat}$ $(h=p/\bar{p})$ LO | NLO | NNLO |
|---|---|---|---|---|---|---|---|---|---|
| BELLE | 0.60 | 0.11 | 0.09 | 0.21 | 0.32 | 0.33 | 0.10 | 0.31 | 0.50 |
| BABAR | 1.91 | 1.77 | 0.78 | 2.86 | 1.11 | 0.95 | 4.74 | 3.75 | 3.25 |
| TASSO12 | 0.70 | 0.85 | 0.87 | 1.10 | 1.03 | 1.02 | 0.69 | 0.70 | 0.72 |
| TASSO14 | 1.55 | 1.67 | 1.70 | 2.17 | 2.13 | 2.07 | 1.32 | 1.25 | 1.22 |
| TASSO22 | 1.64 | 1.91 | 1.91 | 2.14 | 2.77 | 2.62 | 0.98 | 0.92 | 0.93 |
| TPC (incl.) | 0.46 | 0.65 | 0.85 | 0.94 | 1.09 | 1.01 | 1.04 | 1.10 | 1.08 |
| TPC (*uds* tag) | 0.78 | 0.55 | 0.49 | — | — | — | — | — | — |
| TPC (*c* tag) | 0.55 | 0.53 | 0.52 | — | — | — | — | — | — |
| TPC (*b* tag) | 1.44 | 1.43 | 1.43 | — | — | — | — | — | — |
| TASSO30 | — | — | — | — | — | — | 0.25 | 0.19 | 0.18 |
| TASSO34 | 1.16 | 0.98 | 1.00 | 0.27 | 0.44 | 0.36 | 0.82 | 0.81 | 0.78 |
| TASSO44 | 2.01 | 2.24 | 2.34 | — | — | — | — | — | — |
| TOPAZ | 1.04 | 0.82 | 0.80 | 0.61 | 1.19 | 0.99 | 0.79 | 1.21 | 1.19 |
| ALEPH | 1.68 | 0.90 | 0.78 | 0.47 | 0.55 | 0.56 | 1.36 | 1.43 | 1.28 |
| DELPHI (incl.) | 1.44 | 1.79 | 1.86 | 0.28 | 0.33 | 0.34 | 0.48 | 0.49 | 0.49 |
| DELPHI (*uds* tag) | 1.30 | 1.48 | 1.54 | 1.38 | 1.49 | 1.32 | 0.47 | 0.46 | 0.45 |
| DELPHI (*b* tag) | 1.21 | 0.99 | 0.95 | 0.58 | 0.49 | 0.52 | 0.89 | 0.89 | 0.91 |
| OPAL | 2.29 | 1.88 | 1.84 | 1.67 | 1.57 | 1.66 | — | — | — |
| SLD (incl.) | 2.33 | 1.14 | 0.83 | 0.86 | 0.62 | 0.57 | 0.66 | 0.65 | 0.64 |
| SLD (*uds* tag) | 0.95 | 0.65 | 0.52 | 1.31 | 1.02 | 0.93 | 0.77 | 0.76 | 0.78 |
| SLD (*c* tag) | 3.33 | 1.33 | 1.06 | 0.92 | 0.47 | 0.38 | 1.22 | 1.22 | 1.21 |
| SLD (*b* tag) | 0.45 | 0.38 | 0.36 | 0.59 | 0.67 | 0.62 | 1.12 | 1.29 | 1.33 |
| Total dataset | **1.44** | **1.02** | **0.87** | **1.02** | **0.78** | **0.73** | **1.31** | **1.23** | **1.17** |

# Fit quality

- Fit quality **increasingly better** going from LO to NNLO:
  - substantial from LO to NLO, more moderate from NLO to NNLO.
  - **NNLO** corrections are anyway **beneficial** (particularly for pions).
- **Tension** between BELLE and BABAR for kaons and protons:
  - **opposite trend** upon inclusion of higher-order corrections.

| | $\chi^2/N_{\rm dat}$ $(h=\pi^\pm)$ | | | $\chi^2/N_{\rm dat}$ $(h=K^\pm)$ | | | $\chi^2/N_{\rm dat}$ $(h=p/\bar{p})$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Exp. | LO | NLO | NNLO | LO | NLO | NNLO | LO | NLO | NNLO |
| BELLE | 0.60 | 0.11 | 0.09 | 0.21 | 0.32 | 0.33 | 0.10 | 0.31 | 0.50 |
| BABAR | 1.91 | 1.77 | 0.78 | 2.86 | 1.11 | 0.95 | 4.74 | 3.75 | 3.25 |
| TASSO12 | 0.70 | 0.85 | 0.87 | 1.10 | 1.03 | 1.02 | 0.69 | 0.70 | 0.72 |
| TASSO14 | 1.55 | 1.67 | 1.70 | 2.17 | 2.13 | 2.07 | 1.32 | 1.25 | 1.22 |
| TASSO22 | 1.64 | 1.91 | 1.91 | 2.14 | 2.77 | 2.62 | 0.98 | 0.92 | 0.93 |
| TPC (incl.) | 0.46 | 0.65 | 0.85 | 0.94 | 1.09 | 1.01 | 1.04 | 1.10 | 1.08 |
| TPC (*uds* tag) | 0.78 | 0.55 | 0.49 | — | — | — | — | — | — |
| TPC (*c* tag) | 0.55 | 0.53 | 0.52 | — | — | — | — | — | — |
| TPC (*b* tag) | 1.44 | 1.43 | 1.43 | — | — | — | — | — | — |
| TASSO30 | — | — | — | — | — | — | 0.25 | 0.19 | 0.18 |
| TASSO34 | 1.16 | 0.98 | 1.00 | 0.27 | 0.44 | 0.36 | 0.82 | 0.81 | 0.78 |
| TASSO44 | 2.01 | 2.24 | 2.34 | — | — | — | — | — | — |
| TOPAZ | 1.04 | 0.82 | 0.80 | 0.61 | 1.19 | 0.99 | 0.79 | 1.21 | 1.19 |
| ALEPH | 1.68 | 0.90 | 0.78 | 0.47 | 0.55 | 0.56 | 1.36 | 1.43 | 1.28 |
| DELPHI (incl.) | 1.44 | 1.79 | 1.86 | 0.28 | 0.33 | 0.34 | 0.48 | 0.49 | 0.49 |
| DELPHI (*uds* tag) | 1.30 | 1.48 | 1.54 | 1.38 | 1.49 | 1.32 | 0.47 | 0.46 | 0.45 |
| DELPHI (*b* tag) | 1.21 | 0.99 | 0.95 | 0.58 | 0.49 | 0.52 | 0.89 | 0.89 | 0.91 |
| OPAL | 2.29 | 1.88 | 1.84 | 1.67 | 1.57 | 1.66 | — | — | — |
| SLD (incl.) | 2.33 | 1.14 | 0.83 | 0.86 | 0.62 | 0.57 | 0.66 | 0.65 | 0.64 |
| SLD (*uds* tag) | 0.95 | 0.65 | 0.52 | 1.31 | 1.02 | 0.93 | 0.77 | 0.76 | 0.78 |
| SLD (*c* tag) | 3.33 | 1.33 | 1.06 | 0.92 | 0.47 | 0.38 | 1.22 | 1.22 | 1.21 |
| SLD (*b* tag) | 0.45 | 0.38 | 0.36 | 0.59 | 0.67 | 0.62 | 1.12 | 1.29 | 1.33 |
| Total dataset | **1.44** | **1.02** | **0.87** | **1.02** | **0.78** | **0.73** | **1.31** | **1.23** | **1.17** |

# Fit quality

- Fit quality **increasingly better** going from LO to NNLO:
  - substantial from LO to NLO, more moderate from NLO to NNLO.
  - **NNLO** corrections are anyway **beneficial** (particularly for pions).
- **Tension** between BELLE and BABAR for kaons and protons:
  - **opposite trend** upon inclusion of higher-order corrections.
- **Anomalously small** $\chi^2$ for BELLE:
  - possible underestimate of the **uncorrelated systematic** uncertainty.

| | $\chi^2/N_{\rm dat}$ $(h=\pi^\pm)$ | | | $\chi^2/N_{\rm dat}$ $(h=K^\pm)$ | | | $\chi^2/N_{\rm dat}$ $(h=p/\bar{p})$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Exp. | LO | NLO | NNLO | LO | NLO | NNLO | LO | NLO | NNLO |
| BELLE | 0.60 | 0.11 | 0.09 | 0.21 | 0.32 | 0.33 | 0.10 | 0.31 | 0.50 |
| BABAR | 1.91 | 1.77 | 0.78 | 2.86 | 1.11 | 0.95 | 4.74 | 3.75 | 3.25 |
| TASSO12 | 0.70 | 0.85 | 0.87 | 1.10 | 1.03 | 1.02 | 0.69 | 0.70 | 0.72 |
| TASSO14 | 1.55 | 1.67 | 1.70 | 2.17 | 2.13 | 2.07 | 1.32 | 1.25 | 1.22 |
| TASSO22 | 1.64 | 1.91 | 1.91 | 2.14 | 2.77 | 2.62 | 0.98 | 0.92 | 0.93 |
| TPC (incl.) | 0.46 | 0.65 | 0.85 | 0.94 | 1.09 | 1.01 | 1.04 | 1.10 | 1.08 |
| TPC (*uds* tag) | 0.78 | 0.55 | 0.49 | — | — | — | — | — | — |
| TPC (*c* tag) | 0.55 | 0.53 | 0.52 | — | — | — | — | — | — |
| TPC (*b* tag) | 1.44 | 1.43 | 1.43 | — | — | — | — | — | — |
| TASSO30 | — | — | — | — | — | — | 0.25 | 0.19 | 0.18 |
| TASSO34 | 1.16 | 0.98 | 1.00 | 0.27 | 0.44 | 0.36 | 0.82 | 0.81 | 0.78 |
| TASSO44 | 2.01 | 2.24 | 2.34 | — | — | — | — | — | — |
| TOPAZ | 1.04 | 0.82 | 0.80 | 0.61 | 1.19 | 0.99 | 0.79 | 1.21 | 1.19 |
| ALEPH | 1.68 | 0.90 | 0.78 | 0.47 | 0.55 | 0.56 | 1.36 | 1.43 | 1.28 |
| DELPHI (incl.) | 1.44 | 1.79 | 1.86 | 0.28 | 0.33 | 0.34 | 0.48 | 0.49 | 0.49 |
| DELPHI (*uds* tag) | 1.30 | 1.48 | 1.54 | 1.38 | 1.49 | 1.32 | 0.47 | 0.46 | 0.45 |
| DELPHI (*b* tag) | 1.21 | 0.99 | 0.95 | 0.58 | 0.49 | 0.52 | 0.89 | 0.89 | 0.91 |
| OPAL | 2.29 | 1.88 | 1.84 | 1.67 | 1.57 | 1.66 | — | — | — |
| SLD (incl.) | 2.33 | 1.14 | 0.83 | 0.86 | 0.62 | 0.57 | 0.66 | 0.65 | 0.64 |
| SLD (*uds* tag) | 0.95 | 0.65 | 0.52 | 1.31 | 1.02 | 0.93 | 0.77 | 0.76 | 0.78 |
| SLD (*c* tag) | 3.33 | 1.33 | 1.06 | 0.92 | 0.47 | 0.38 | 1.22 | 1.22 | 1.21 |
| SLD (*b* tag) | 0.45 | 0.38 | 0.36 | 0.59 | 0.67 | 0.62 | 1.12 | 1.29 | 1.33 |
| Total dataset | **1.44** | **1.02** | **0.87** | **1.02** | **0.78** | **0.73** | **1.31** | **1.23** | **1.17** |

# Fit quality

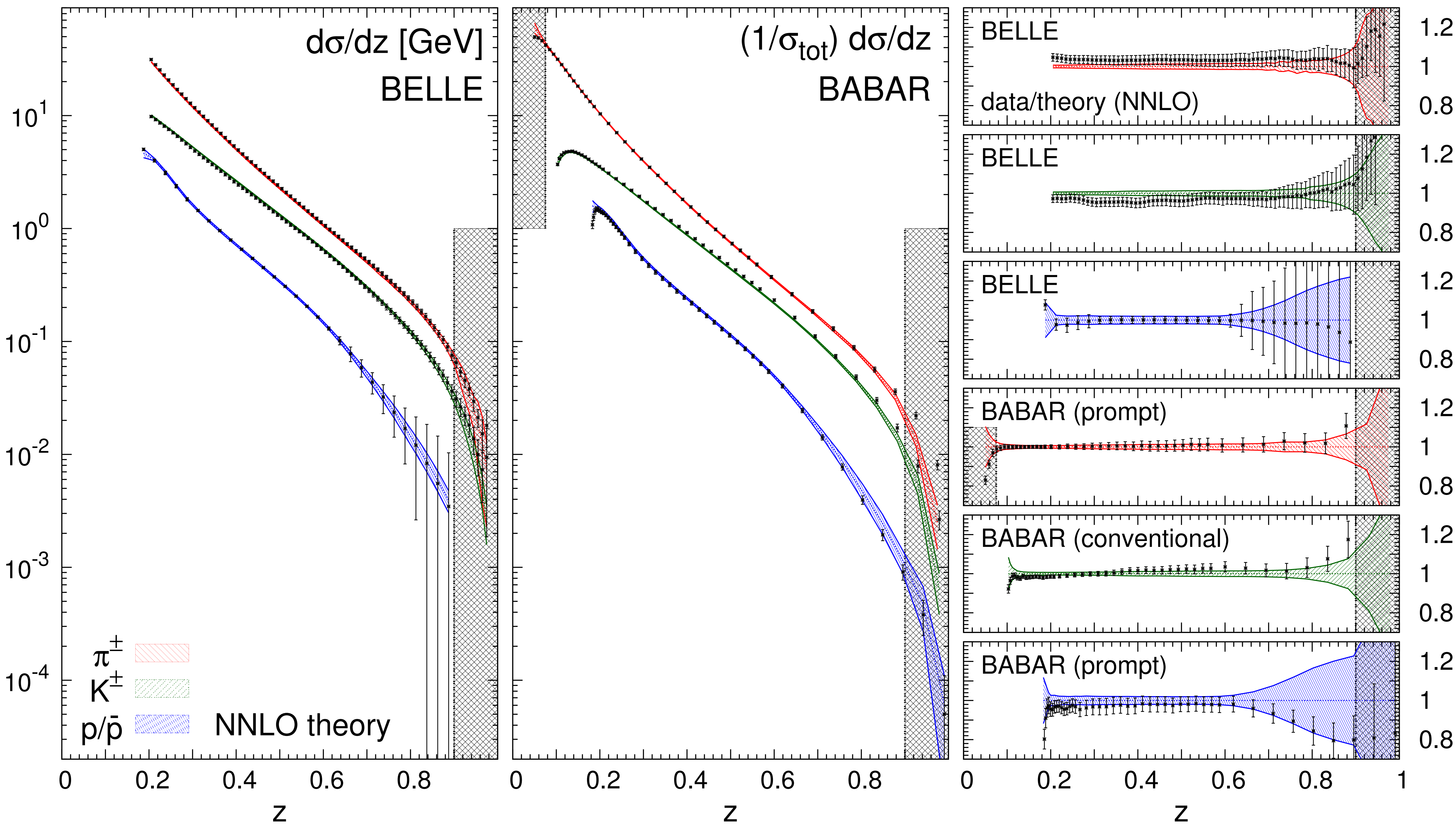
- Fit quality **increasingly better** going from LO to NNLO:
  - substantial from LO to NLO, more moderate from NLO to NNLO.
  - **NNLO** corrections are anyway **beneficial** (particularly for pions).
- **Tension** between BELLE and BABAR for kaons and protons:
  - **opposite trend** upon inclusion of higher-order corrections.
- **Anomalously small** $\chi^2$ for BELLE:
  - possible underestimate of the **uncorrelated systematic** uncertainty.

| Exp. | $\chi^2/N_{\rm dat}$ $(h=\pi^\pm)$ LO | NLO | NNLO | $\chi^2/N_{\rm dat}$ $(h=K^\pm)$ LO | NLO | NNLO | $\chi^2/N_{\rm dat}$ $(h=p/\bar{p})$ LO | NLO | NNLO |
|---|---|---|---|---|---|---|---|---|---|
| BELLE | 0.60 | 0.11 | 0.09 | 0.21 | 0.32 | 0.33 | 0.10 | 0.31 | 0.50 |
| BABAR | 1.91 | 1.77 | 0.78 | 2.86 | 1.11 | 0.95 | 4.74 | 3.75 | 3.25 |
| TASSO12 | 0.70 | 0.85 | 0.87 | 1.10 | 1.03 | 1.02 | 0.69 | 0.70 | 0.72 |
| TASSO14 | 1.55 | 1.67 | 1.70 | 2.17 | 2.13 | 2.07 | 1.32 | 1.25 | 1.22 |
| TASSO22 | 1.64 | 1.91 | 1.91 | 2.14 | 2.77 | 2.62 | 0.98 | 0.92 | 0.93 |
| TPC (incl.) | 0.46 | 0.65 | 0.85 | 0.94 | 1.09 | 1.01 | 1.04 | 1.10 | 1.08 |
| TPC (*uds* tag) | 0.78 | 0.55 | 0.49 | — | — | — | — | — | — |
| TPC (*c* tag) | 0.55 | 0.53 | 0.52 | — | — | — | — | — | — |
| TPC (*b* tag) | 1.44 | 1.43 | 1.43 | — | — | — | — | — | — |
| TASSO30 | — | — | — | — | — | — | 0.25 | 0.19 | 0.18 |
| TASSO34 | 1.16 | 0.98 | 1.00 | 0.27 | 0.44 | 0.36 | 0.82 | 0.81 | 0.78 |
| TASSO44 | 2.01 | 2.24 | 2.34 | — | — | — | — | — | — |
| TOPAZ | 1.04 | 0.82 | 0.80 | 0.61 | 1.19 | 0.99 | 0.79 | 1.21 | 1.19 |
| ALEPH | 1.68 | 0.90 | 0.78 | 0.47 | 0.55 | 0.56 | 1.36 | 1.43 | 1.28 |
| DELPHI (incl.) | 1.44 | 1.79 | 1.86 | 0.28 | 0.33 | 0.34 | 0.48 | 0.49 | 0.49 |
| DELPHI (*uds* tag) | 1.30 | 1.48 | 1.54 | 1.38 | 1.49 | 1.32 | 0.47 | 0.46 | 0.45 |
| DELPHI (*b* tag) | 1.21 | 0.99 | 0.95 | 0.58 | 0.49 | 0.52 | 0.89 | 0.89 | 0.91 |
| OPAL | 2.29 | 1.88 | 1.84 | 1.67 | 1.57 | 1.66 | — | — | — |
| SLD (incl.) | 2.33 | 1.14 | 0.83 | 0.86 | 0.62 | 0.57 | 0.66 | 0.65 | 0.64 |
| SLD (*uds* tag) | 0.95 | 0.65 | 0.52 | 1.31 | 1.02 | 0.93 | 0.77 | 0.76 | 0.78 |
| SLD (*c* tag) | 3.33 | 1.33 | 1.06 | 0.92 | 0.47 | 0.38 | 1.22 | 1.22 | 1.21 |
| SLD (*b* tag) | 0.45 | 0.38 | 0.36 | 0.59 | 0.67 | 0.62 | 1.12 | 1.29 | 1.33 |
| Total dataset | **1.44** | **1.02** | **0.87** | **1.02** | **0.78** | **0.73** | **1.31** | **1.23** | **1.17** |

- **Possible tension** also between DELPHI inclusive and the other experiments at $M_Z$:
  - opposite trend upon inclusion of higher-order corrections.

# Description of the data

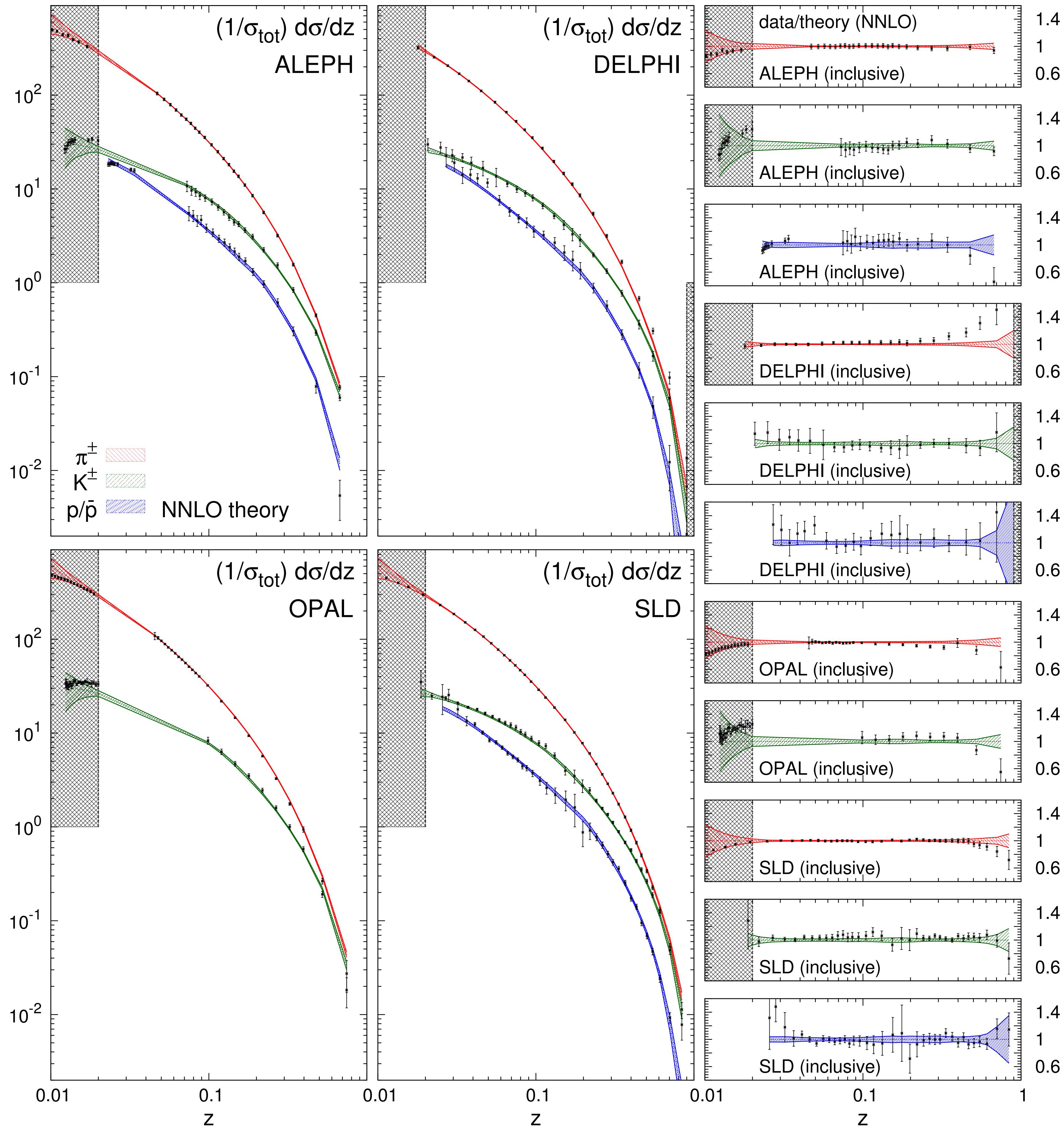
- Data/Theory comparison for **BELLE** and **BABAR** using NNFF1.0 at NNLO:
  - the bands indicate the 1-σ uncertainty.




- Very good description in the region not excluded by the kinematic cuts (shaded areas).
- Different **trend** of the data at **low *z*** for **kaons** and particularly for **protons**:
  - possible reason of the worsening of the $\chi^2$.

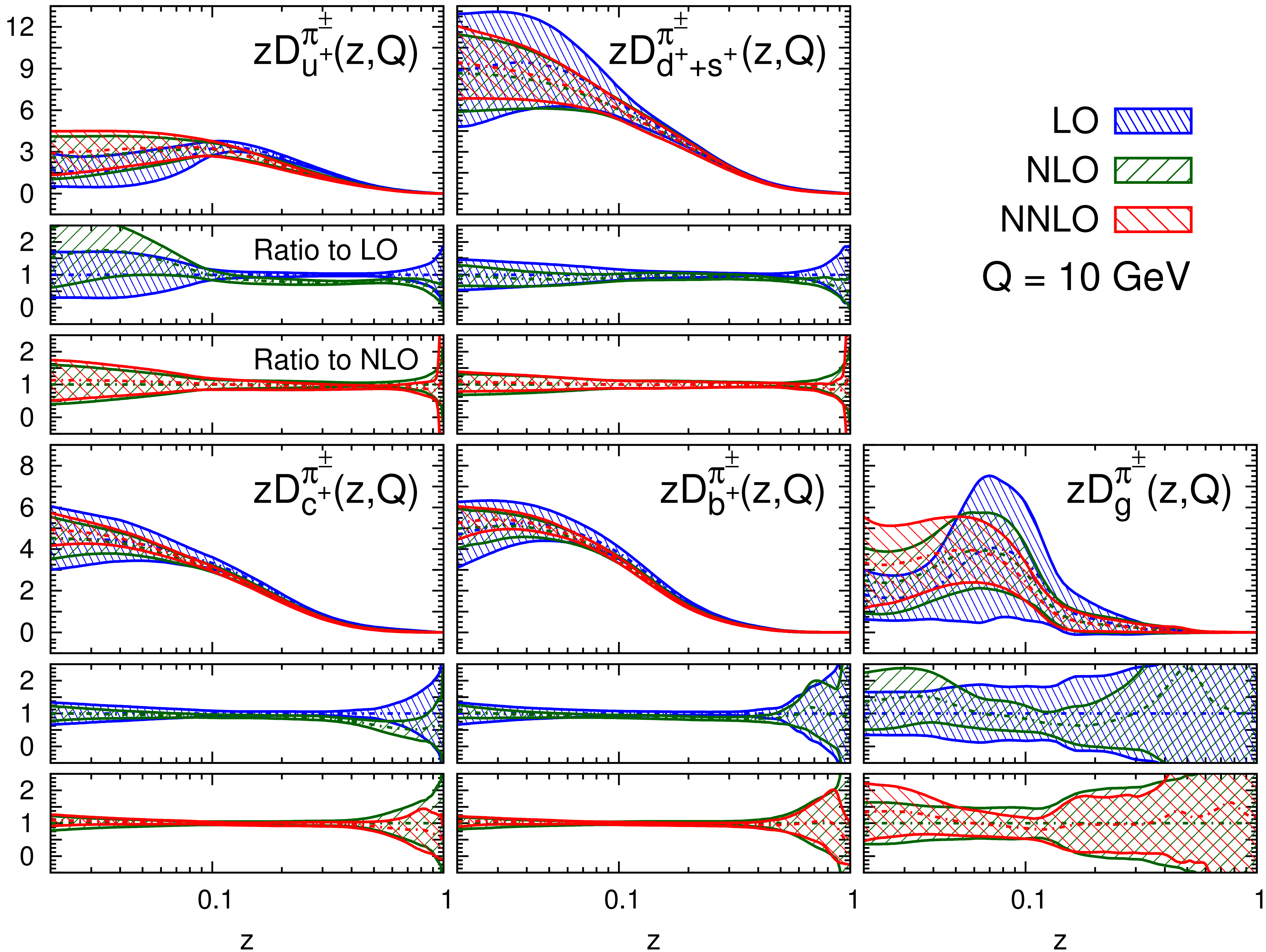
# Description of the data

- Data/Theory comparison for the experiments at $M_Z$ using NNFF1.0 at NNLO.
- Very good description in the region allowed by the kinematic cuts.
- Often also the data excluded by the cuts are well described.
- The predictions for pions for **DELPHI** overshoot the data:
    - origin of the worse $\chi^2$ as compared to the other experiments at $M_Z$.



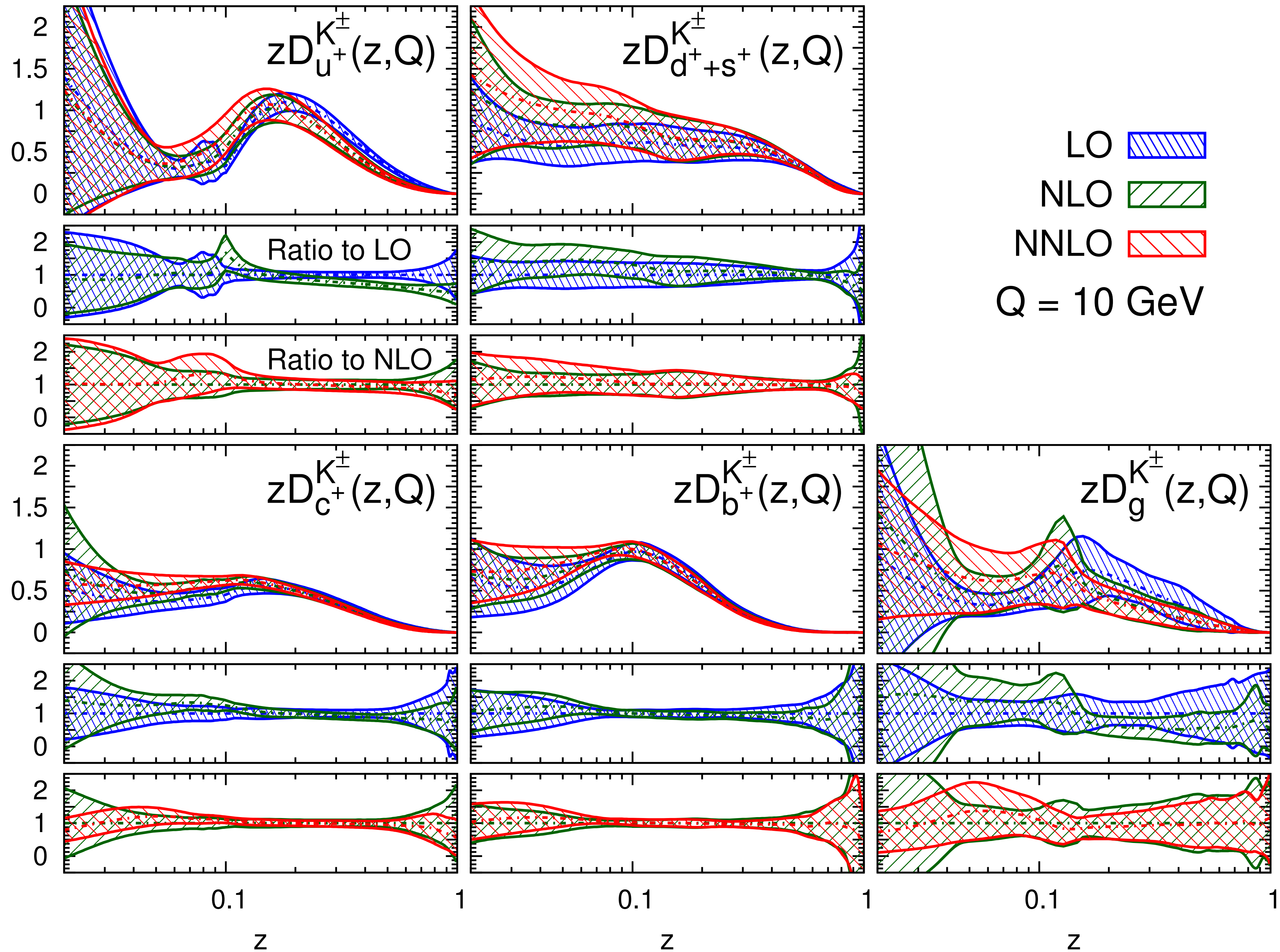
# Fragmentation functions

*Perturbative stability (Pions)*




- **Stabilisation** going from LO to NNLO,
- LO uncertainties slightly larger: poorer theoretical description.

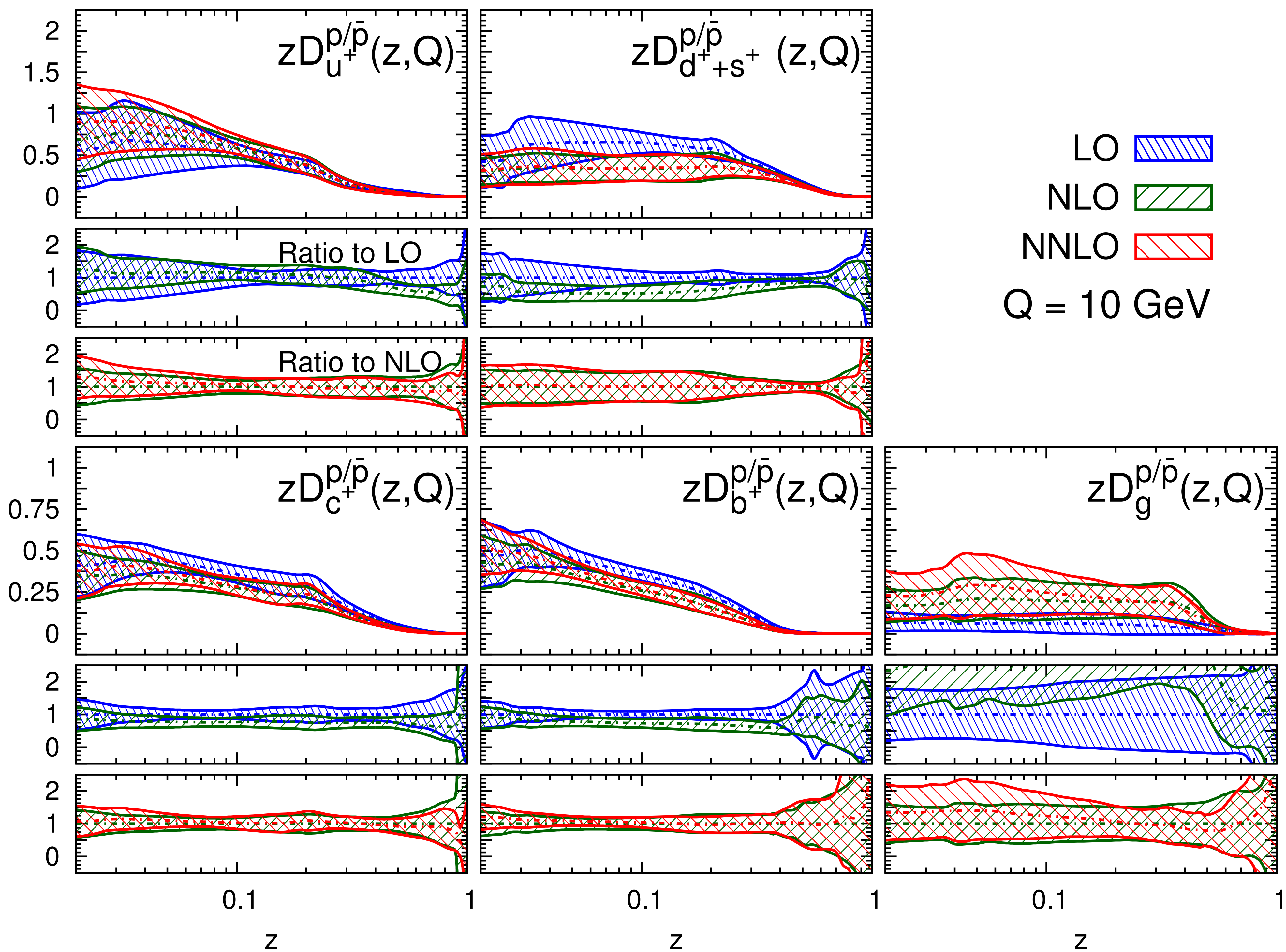
# Fragmentation functions
## *Perturbative stability (Kaons)*




- Same for kaons…

# Fragmentation functions

*Perturbative stability (Protons)*




- …and for protons.

# Fitting methodology

## *The NNPDF approach*

- In the NNPDF procedure applied to PDFs the parametrisation is:

$$f_i(x) = \boxed{x^{\alpha_i}(1-x)^{\beta_i}} \mathrm{NN}_i(x)$$

Preprocessing function

- The preprocessing function:
  - helps implement **physical constraints** (*e.g.* $f_i(1) = 0$ and integrability),
  - determines the behaviour in the **extrapolation regions**,
  - facilitates the task of the neural network making the **fit easier**.
- The values of $\alpha_i$ and $\beta_i$ are **iteratively** determined from data.
- For the fits of FFs we remove the preprocessing functions and use:

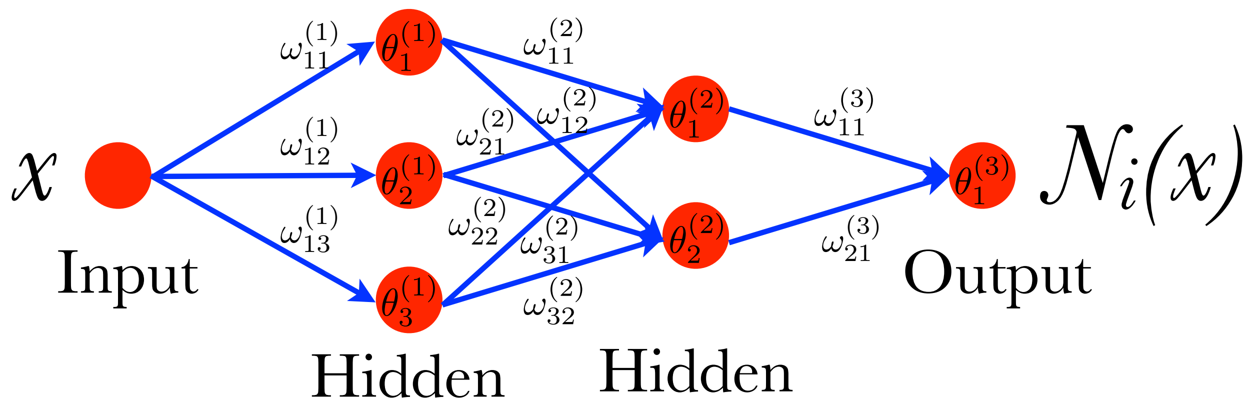
$$f_i(x) = \mathrm{NN}_i(x) - \mathrm{NN}_i(1)$$

- **no need to iterate** to determine $\alpha_i$ and $\beta_i$.
- the NN defines the behaviour also in the extrapolation regions.

# Fitting methodology

*The NNPDF approach*

- **Removing the preprocessing** functions requires a proper choice of the **activation function** of the neural networks:



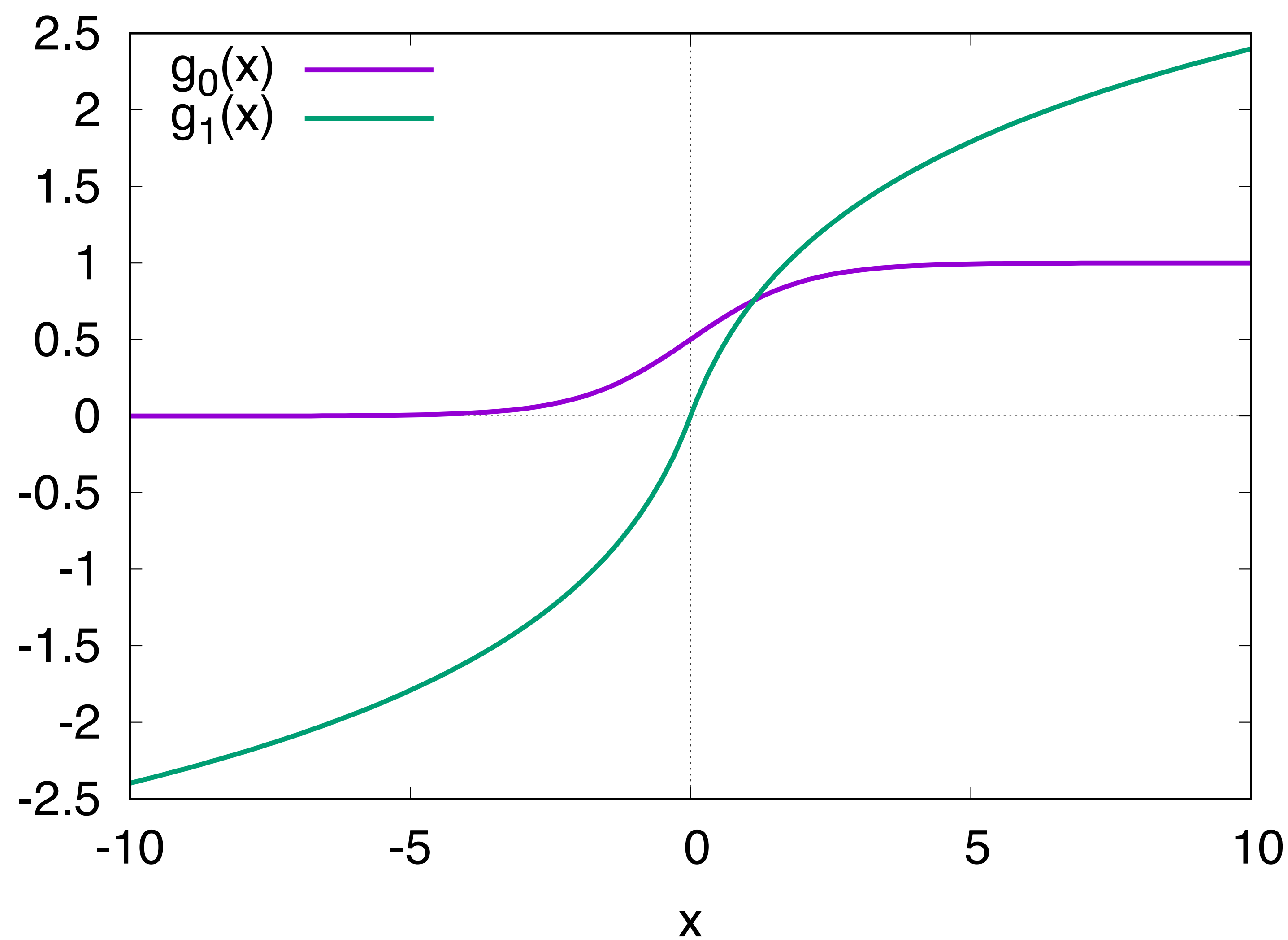

$$\xi_i^{(j)} = g\left(\sum_k^{\text{(j-1)th layer}} \xi_k^{(j-1)} \omega_{ki}^{(j)} - \theta_i^{(j)}\right)$$

Saturating function

$$g_0(x) = \frac{1}{1+e^{-x}}$$

Non-saturating function

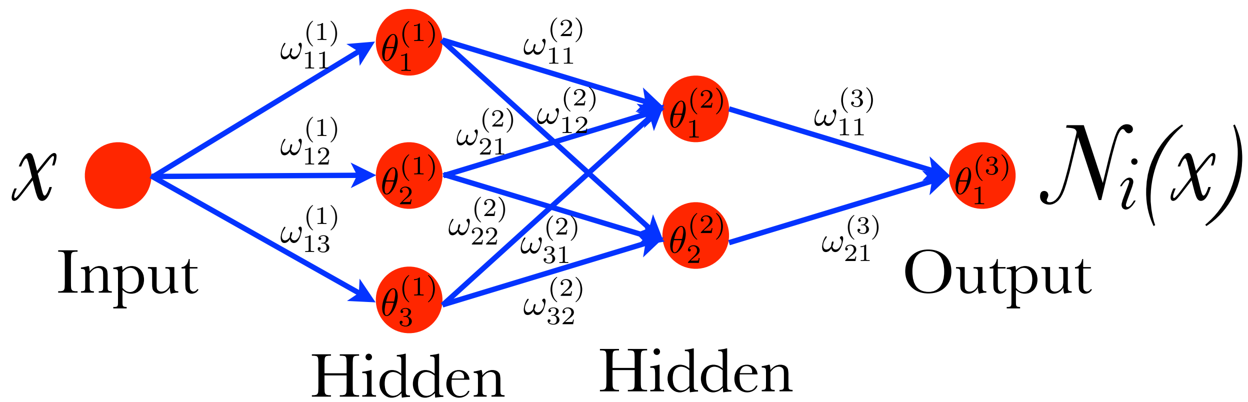
$$g_1(x) = \mathrm{sign}(x)\ln(|x|+1)$$



# Fitting methodology

## *The NNPDF approach*

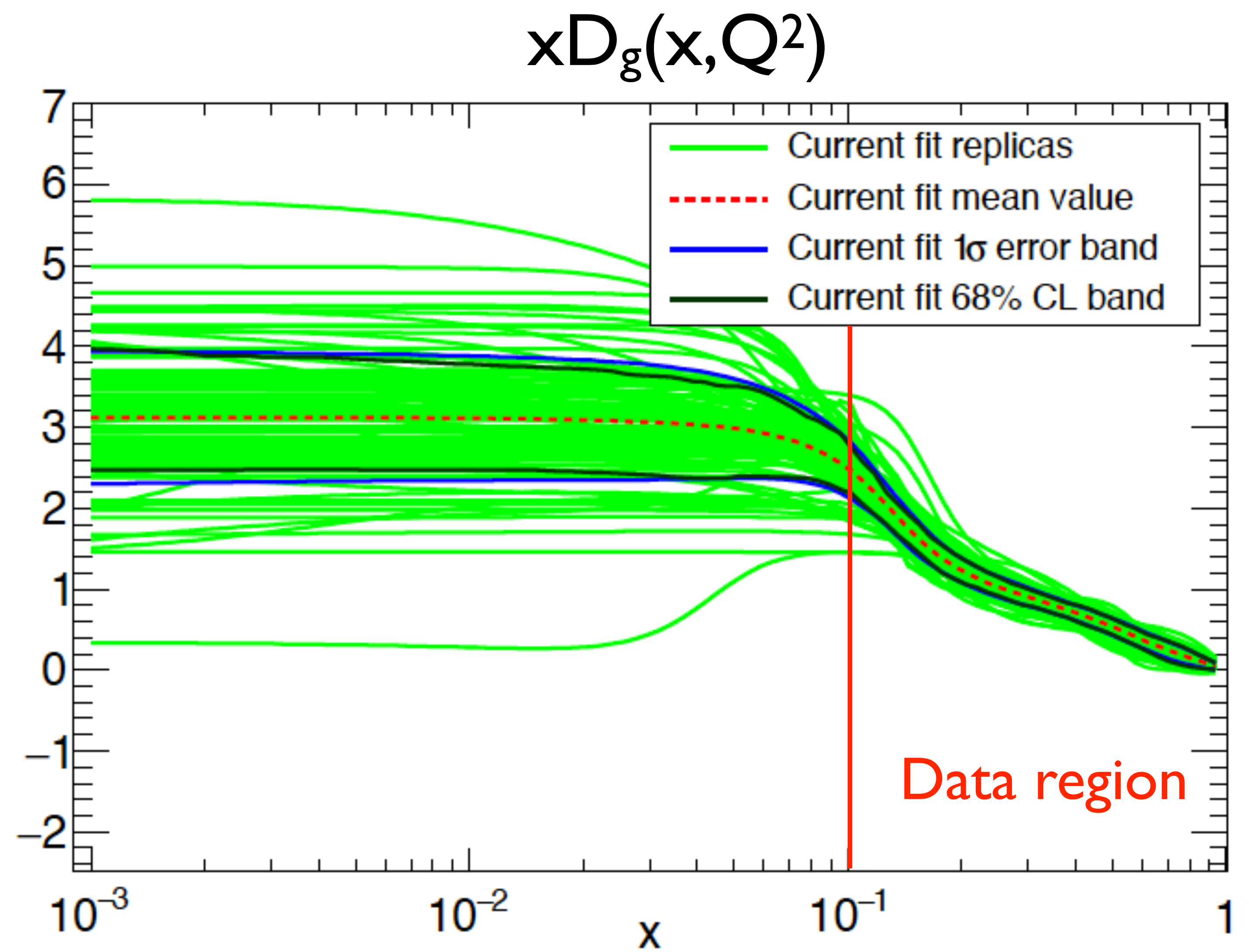
- **Removing the preprocessing** functions requires a proper choice of the **activation function** of the neural networks:




$$\xi_i^{(j)} = g\left(\sum_k^{\text{(j-1)th layer}} \xi_k^{(j-1)} \omega_{ki}^{(j)} - \theta_i^{(j)}\right)$$

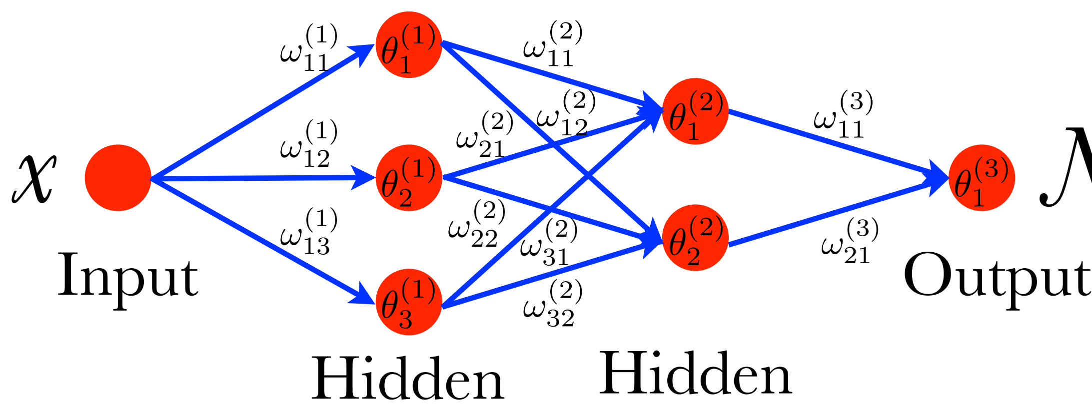
Saturating function

$$g_0(x) = \frac{1}{1+e^{-x}}$$



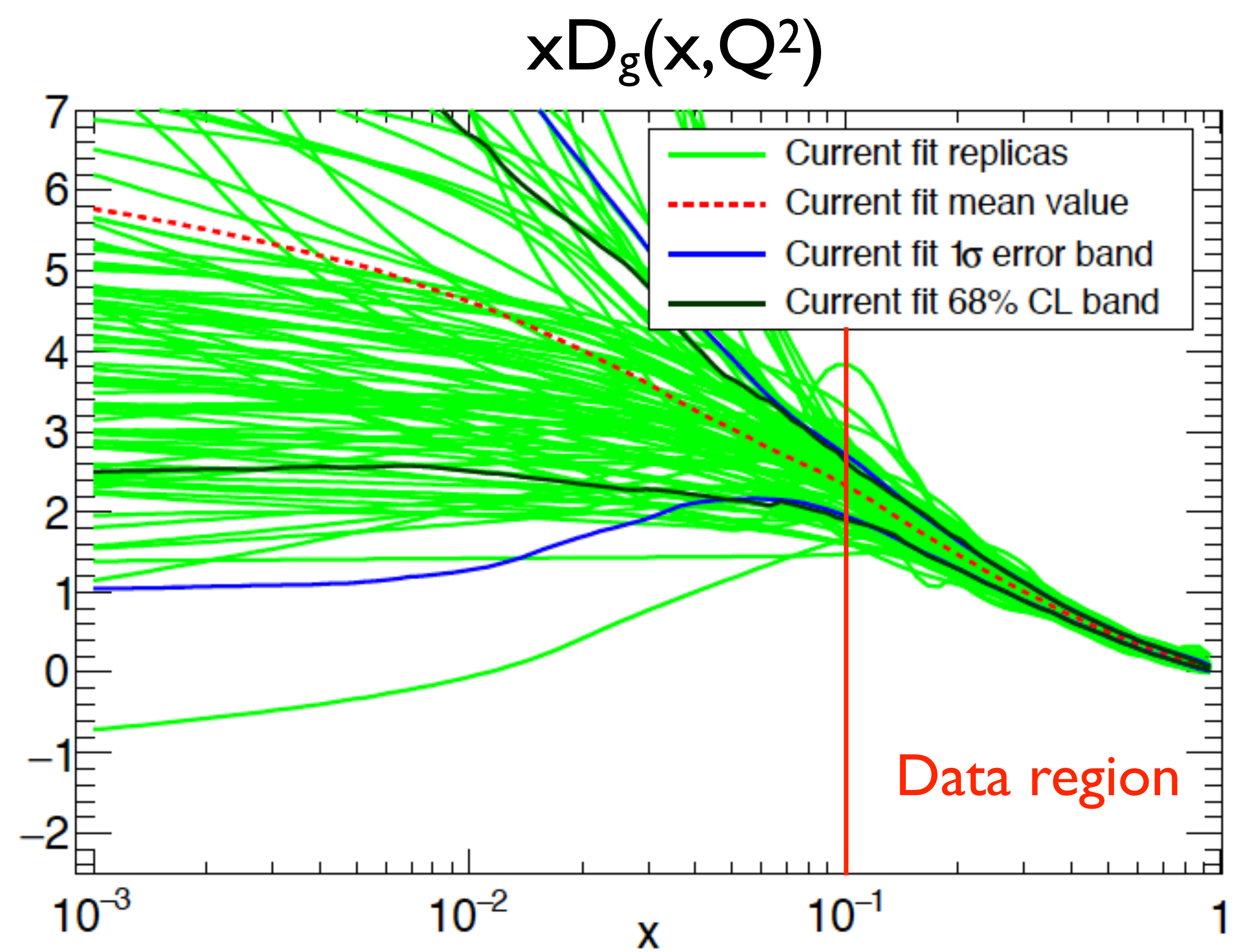
# Fitting methodology
## *The NNPDF approach*

- **Removing the preprocessing** functions requires a proper choice of the **activation function** of the neural networks:




$\mathcal{N}_i(x)$

$$\xi_i^{(j)} = g\left( \sum_k^{\text{(j-1)th layer}} \xi_k^{(j-1)} \omega_{ki}^{(j)} - \theta_i^{(j)} \right)$$

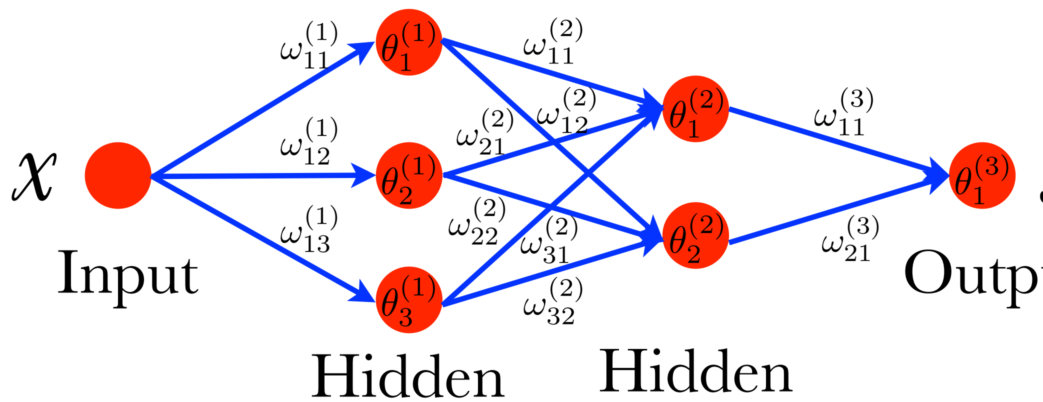



Non-saturating function

$$g_1(x) = \mathrm{sign}(x)\ln(|x|+1)$$

# Fitting methodology
*The NNPDF approach*

- **Removing the preprocessing** functions requires a proper choice of the **activation function** of the neural networks:

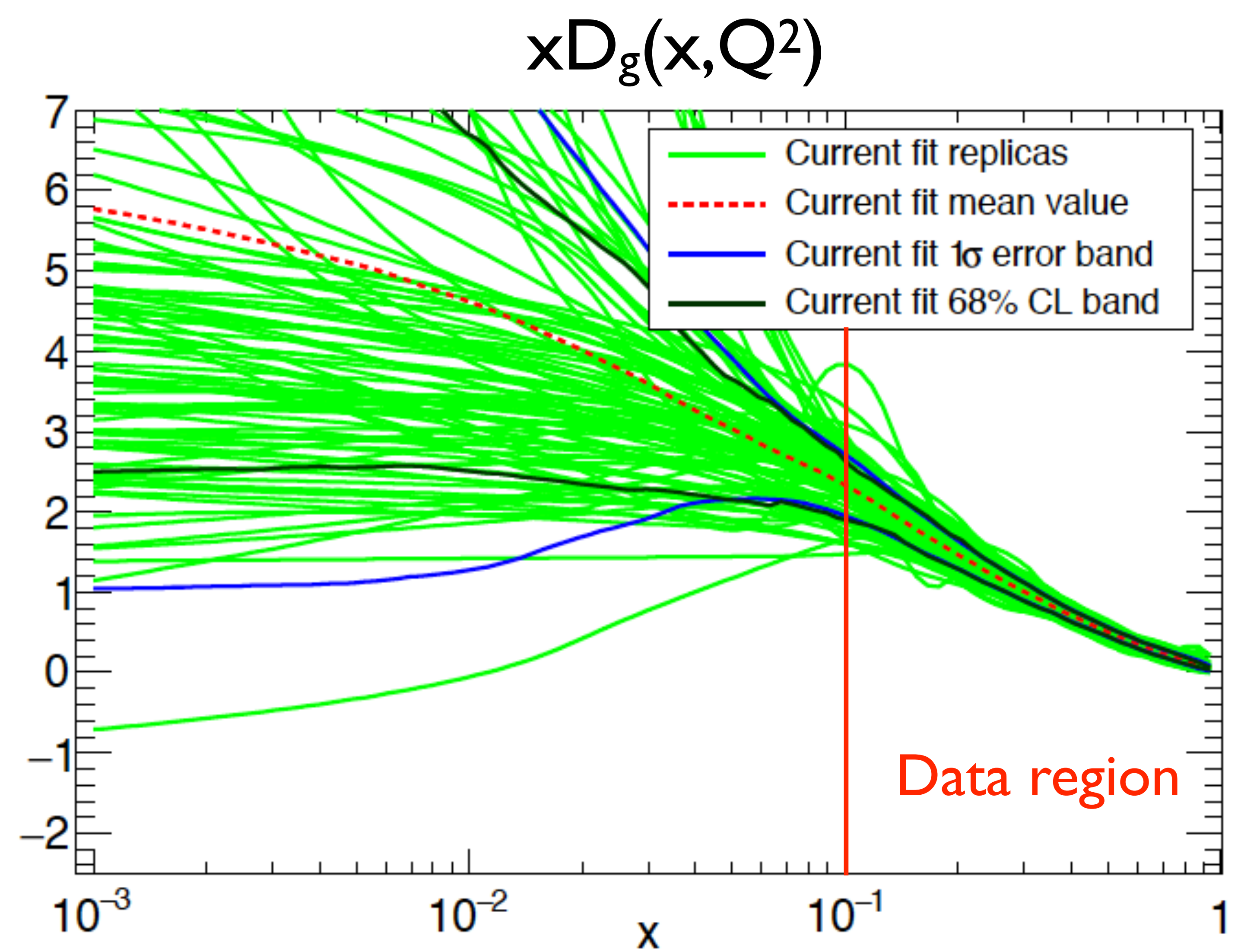



$\mathcal{N}_i(x)$ $\quad \xi_i^{(j)} = g\left(\sum_k^{\text{(j-1)th layer}} \xi_k^{(j-1)} \omega_{ki}^{(j)} - \theta_i^{(j)}\right)$

Non-saturating function

$$g_1(x) = \mathrm{sign}(x)\ln(|x|+1)$$

Our choice for the FFs



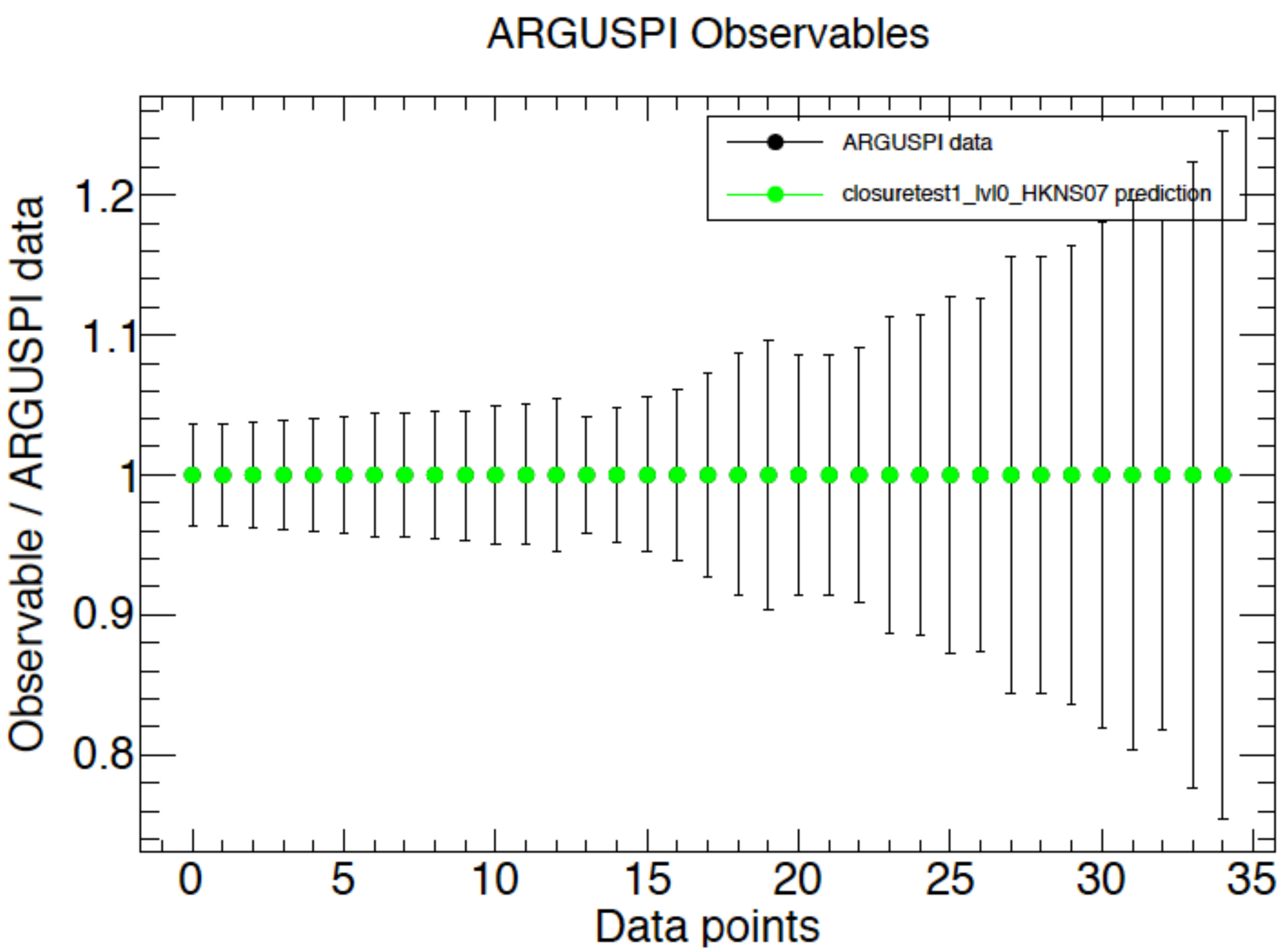
# Closure tests

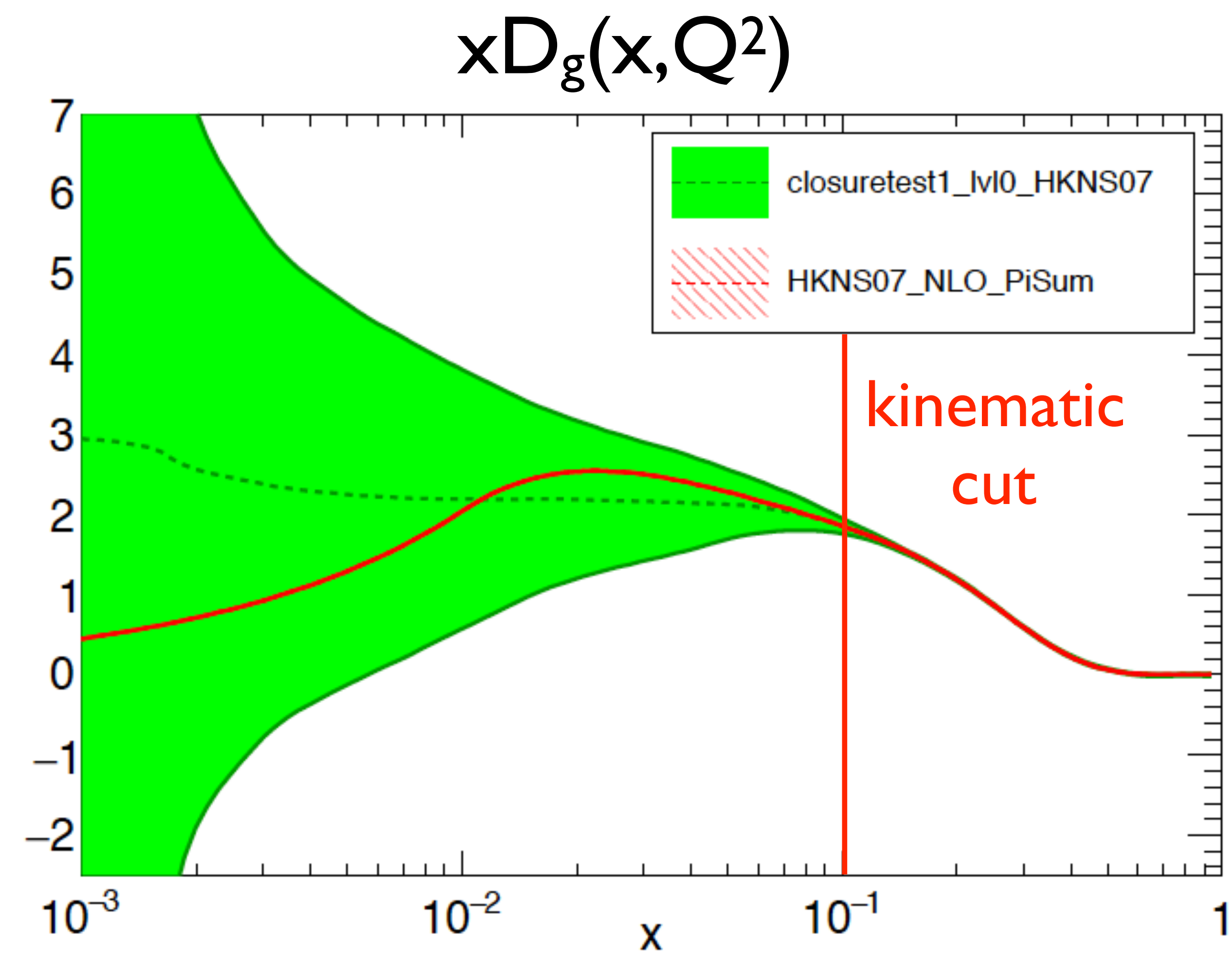
- How do we know whether our fitting strategy is reliable?
  1) **Assume underlying FFs** are known (*e.g.* HKNS07).
  2) **Generate pseudo-data** with given statistical and correlated systematics.
  3) **Perform a fit** and compare to the "truth".
- If needed, use the closure tests to **tune** the fitting algorithm.
- Levels of closure tests: NNPDF Collaboration [arXiv:1410.8849]
  - **level 0**:
    - data point central values equal to the HKNS07 "true" values,
    - uncertainties assumed equal to the experimental ones,
    - we must find $\boldsymbol{\chi^2 \sim 0}$ and that **uncertainty on predictions tends to zero**.
  - **level 2**:
    - data points obtained as random fluctuations with exp. covariance matrix about the "truth",
    - generate Monte Carlo replicas of this data,
    - fit a PDF set to each Monte Carlo replica,
    - we must find $\boldsymbol{\chi^2 \sim 1}$ and that HKNS07 "true" FFs are **within the 1-σ band**.

# Closure tests: "level 0"

- We find:

$$\chi^2 \text{ (global)} = 0.00027$$



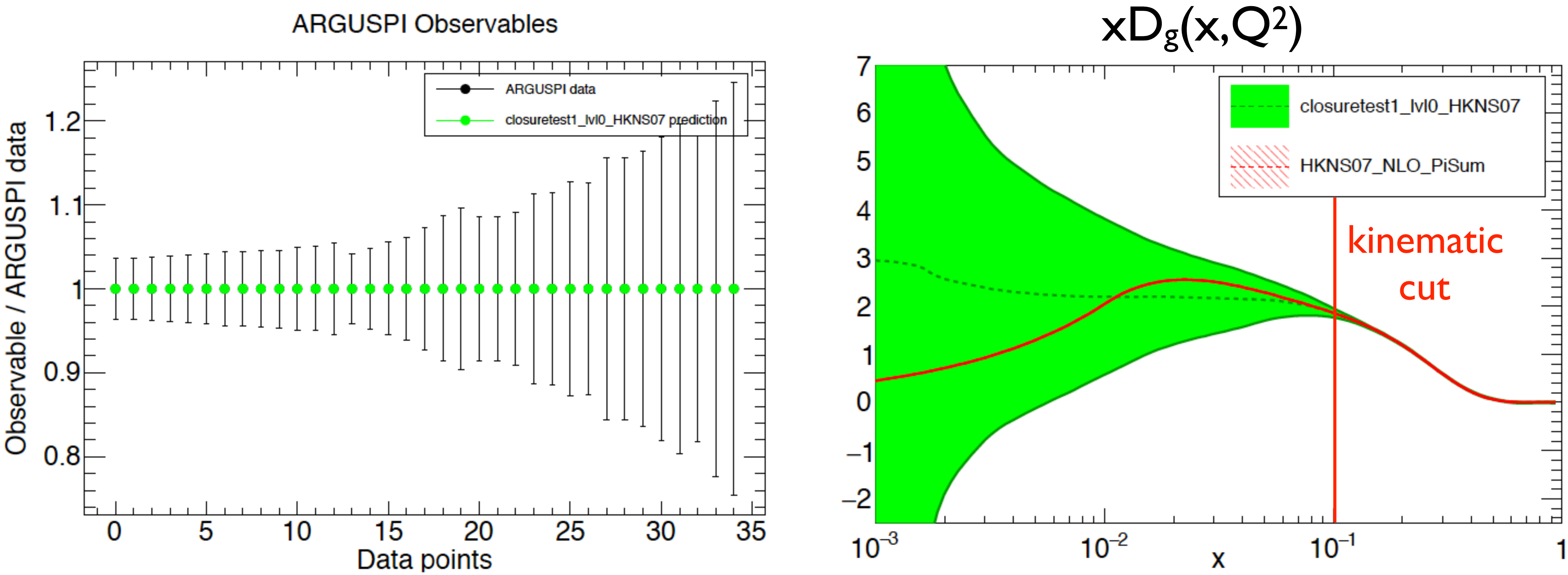




- Predictions coincide with the data central values.
- Prediction uncertainties shrink to zero.
- FFs in the data region very close to the "truth".
- Uncertainties blow up in the extrapolation region.

# Closure tests: “level 0”

- We find:

$$\chi^2 \text{ (global)} = 0.00027$$



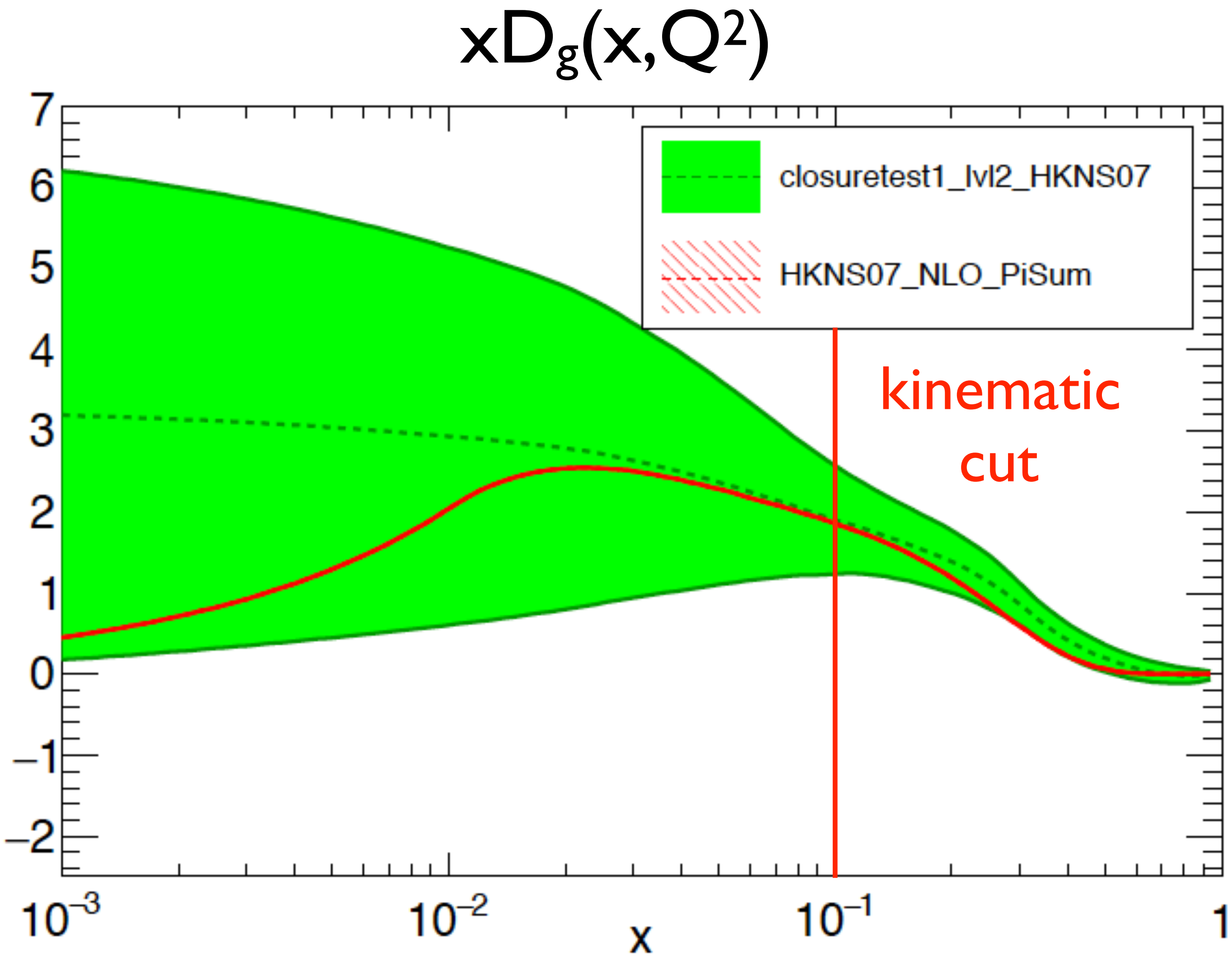

- Predictions coincide with the data central values.
- Prediction uncertainties shrink to zero.
- FFs in the data region very close to the “truth”.
- Uncertainties blow up in the extrapolation region.

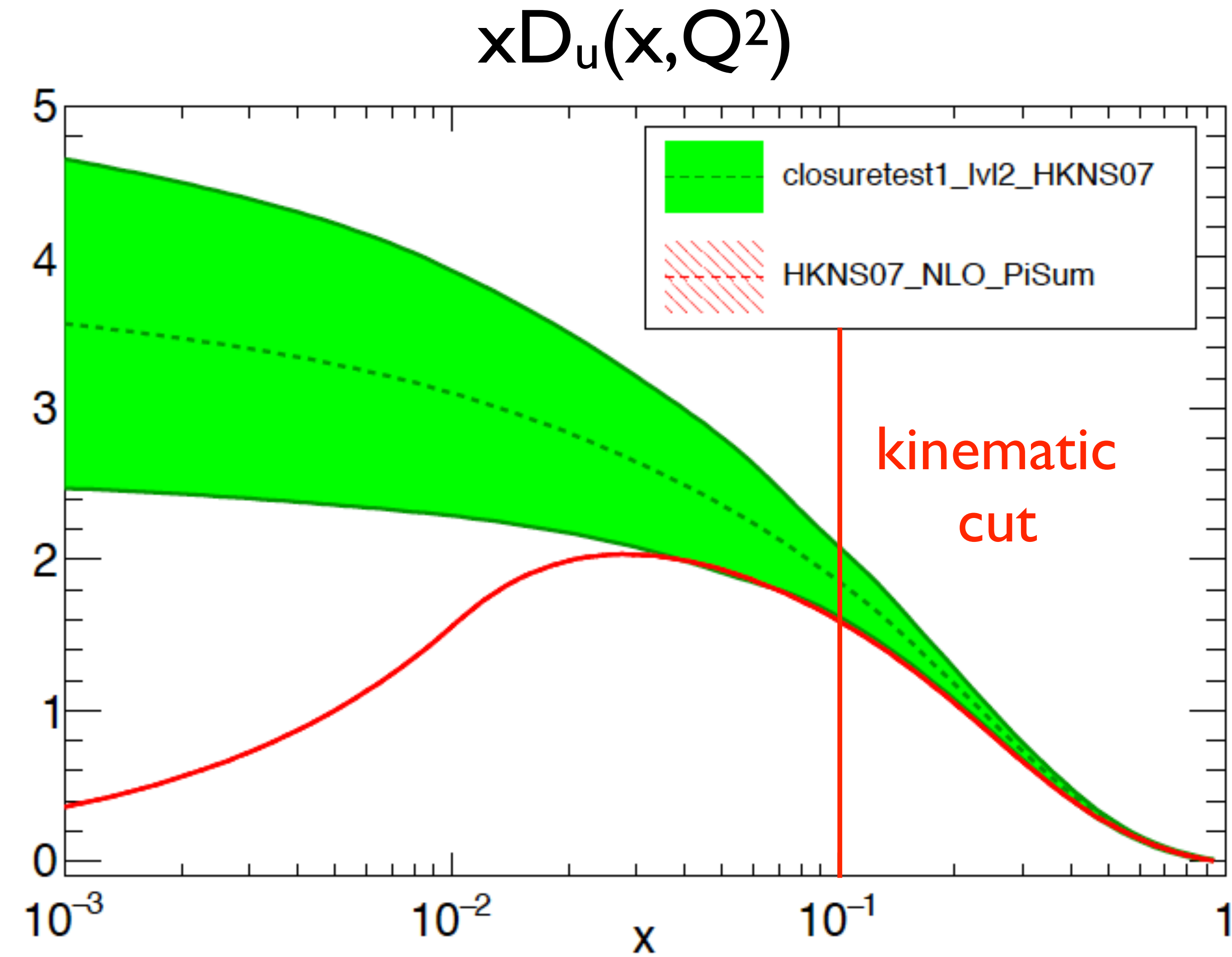
Functional uncertainty

# Closure tests: "level 2"

- We find:

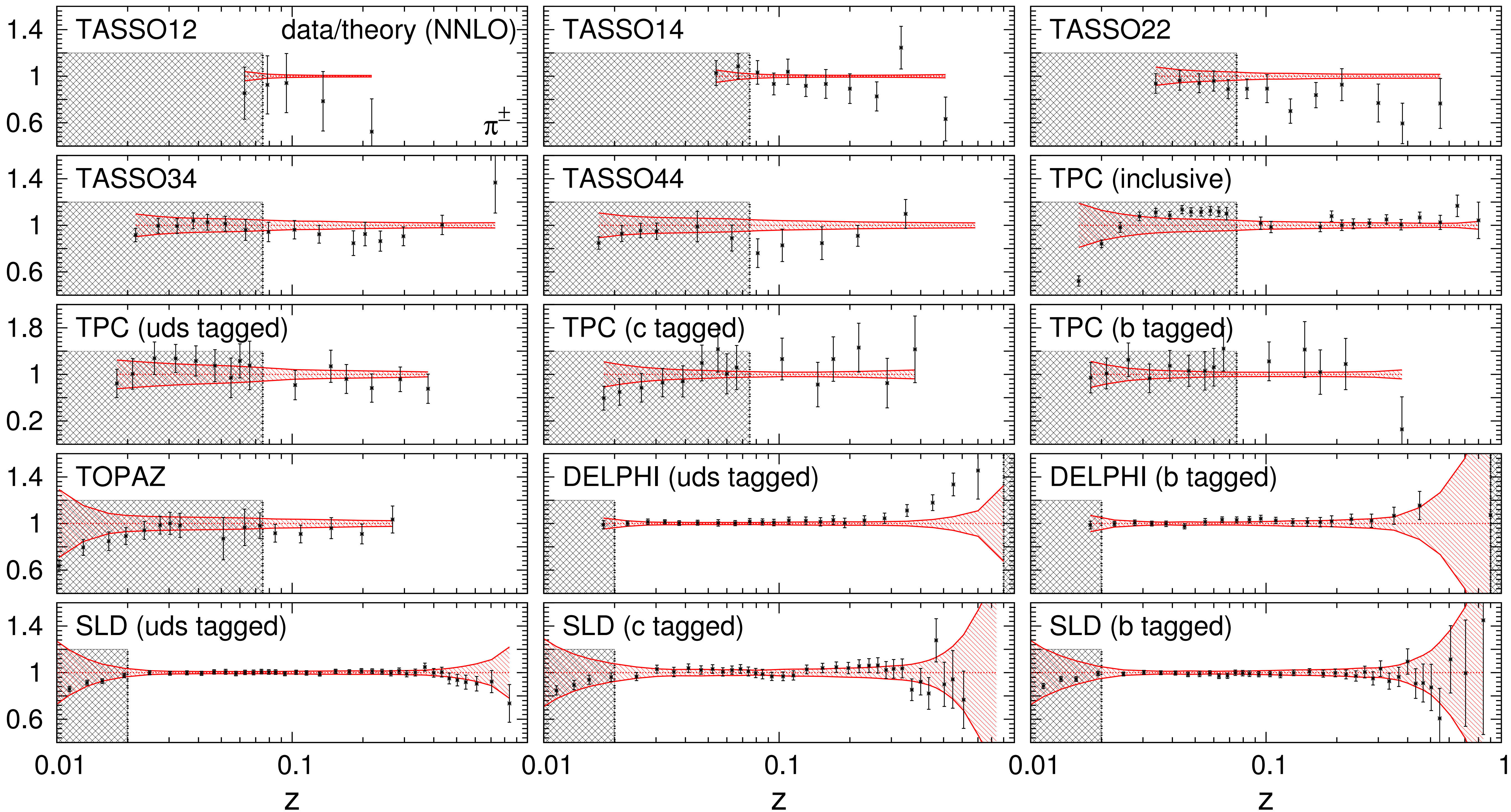
$$\chi^2 \text{ (global)} = 0.99307$$



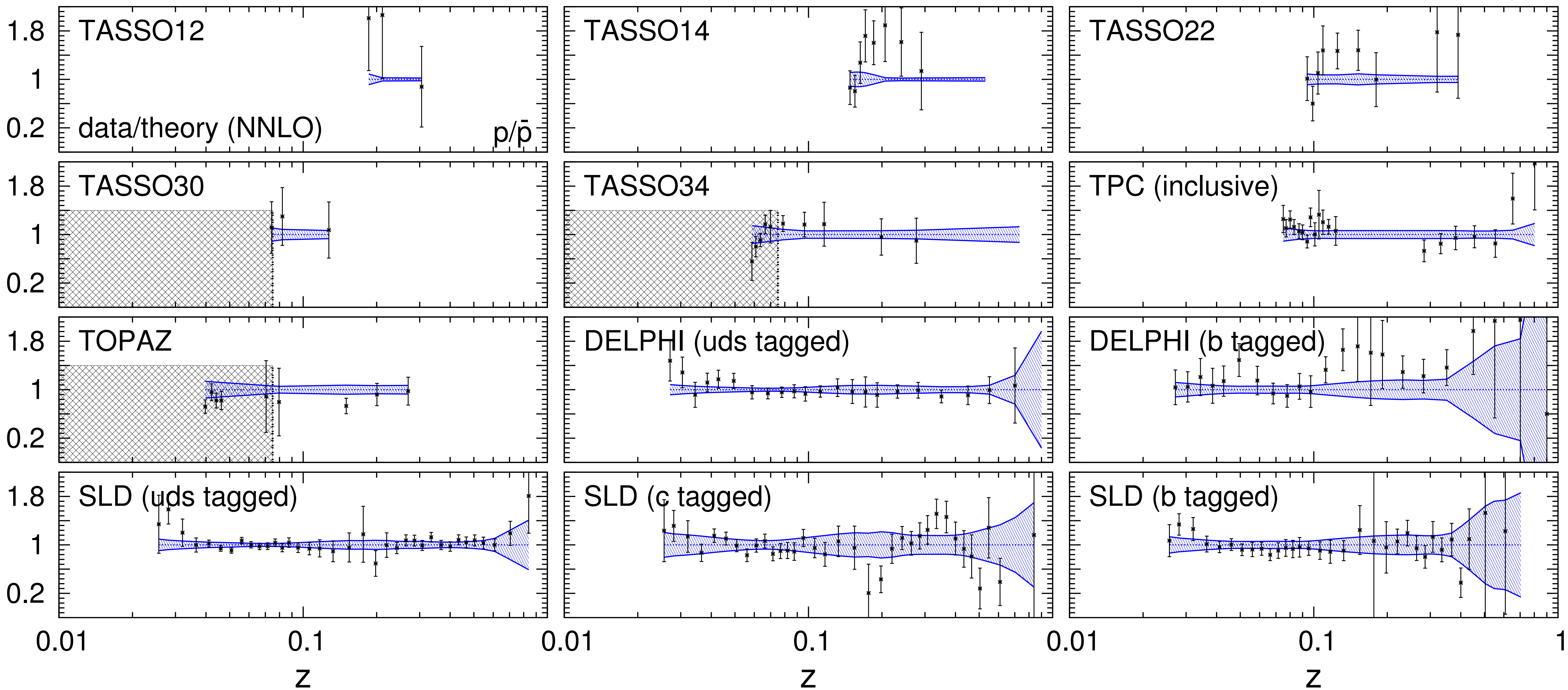



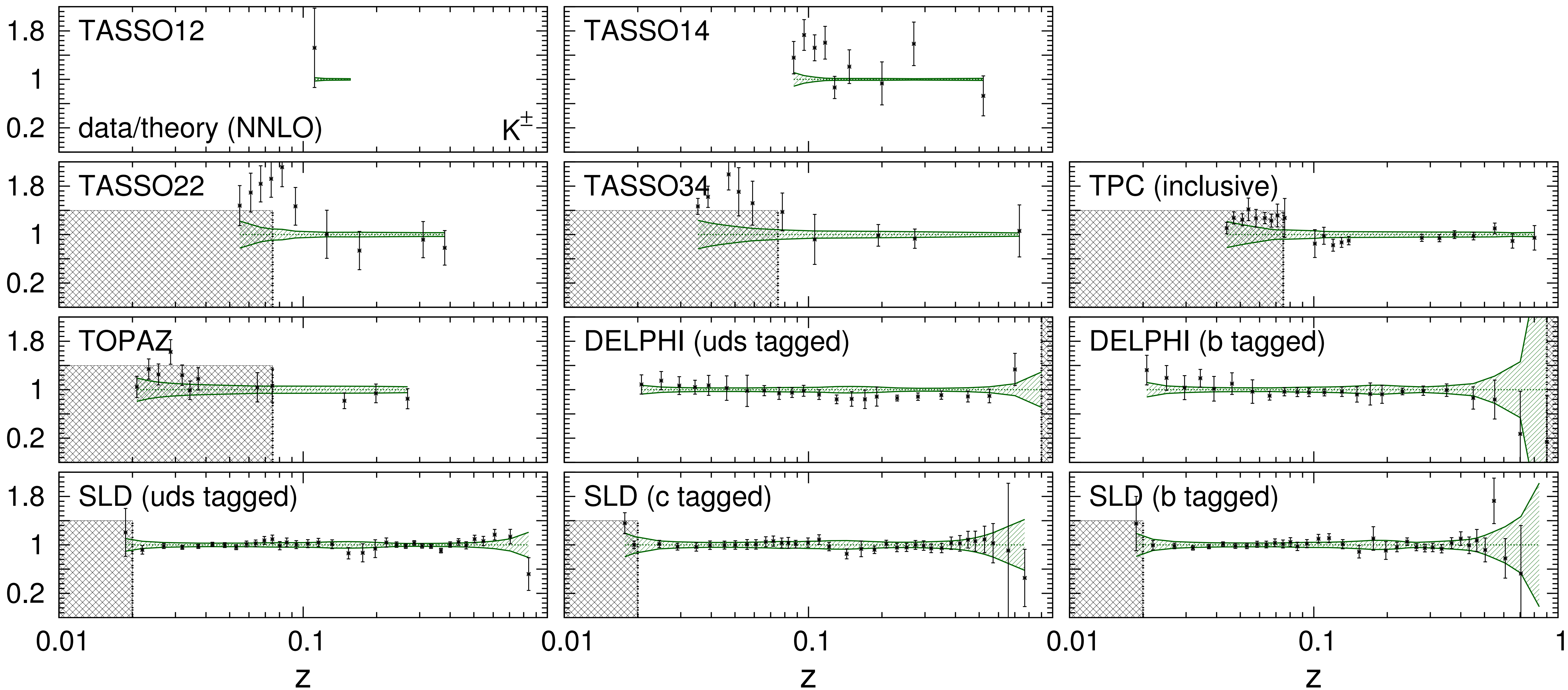

- "True" FFs do fall within the 1-σ band of the fitted FFs (in the data region),

- **Faithful representation** of the experimental uncertainty.



TASSO12
data/theory (NNLO)
$\pi^{\pm}$
TASSO14
TASSO22
TASSO34
TASSO44
TPC (inclusive)
TPC (uds tagged)
TPC (c tagged)
TPC (b tagged)
TOPAZ
DELPHI (uds tagged)
DELPHI (b tagged)
SLD (uds tagged)
SLD (c tagged)
SLD (b tagged)
1.4
1
0.6
1.8
0.2
0.01
0.1
1
z


TASSO12
data/theory (NNLO)
p/p̄
TASSO14
TASSO22
TASSO30
TASSO34
TPC (inclusive)
TOPAZ
DELPHI (uds tagged)
DELPHI (b tagged)
SLD (uds tagged)
SLD (c tagged)
SLD (b tagged)
1.8
1
0.2
0.01
0.1
1
z


TASSO12
TASSO14
data/theory (NNLO)
$K^{\pm}$
TASSO22
TASSO34
TPC (inclusive)
TOPAZ
DELPHI (uds tagged)
DELPHI (b tagged)
SLD (uds tagged)
SLD (c tagged)
SLD (b tagged)
1.8
1
0.2
0.01
0.1
1
z